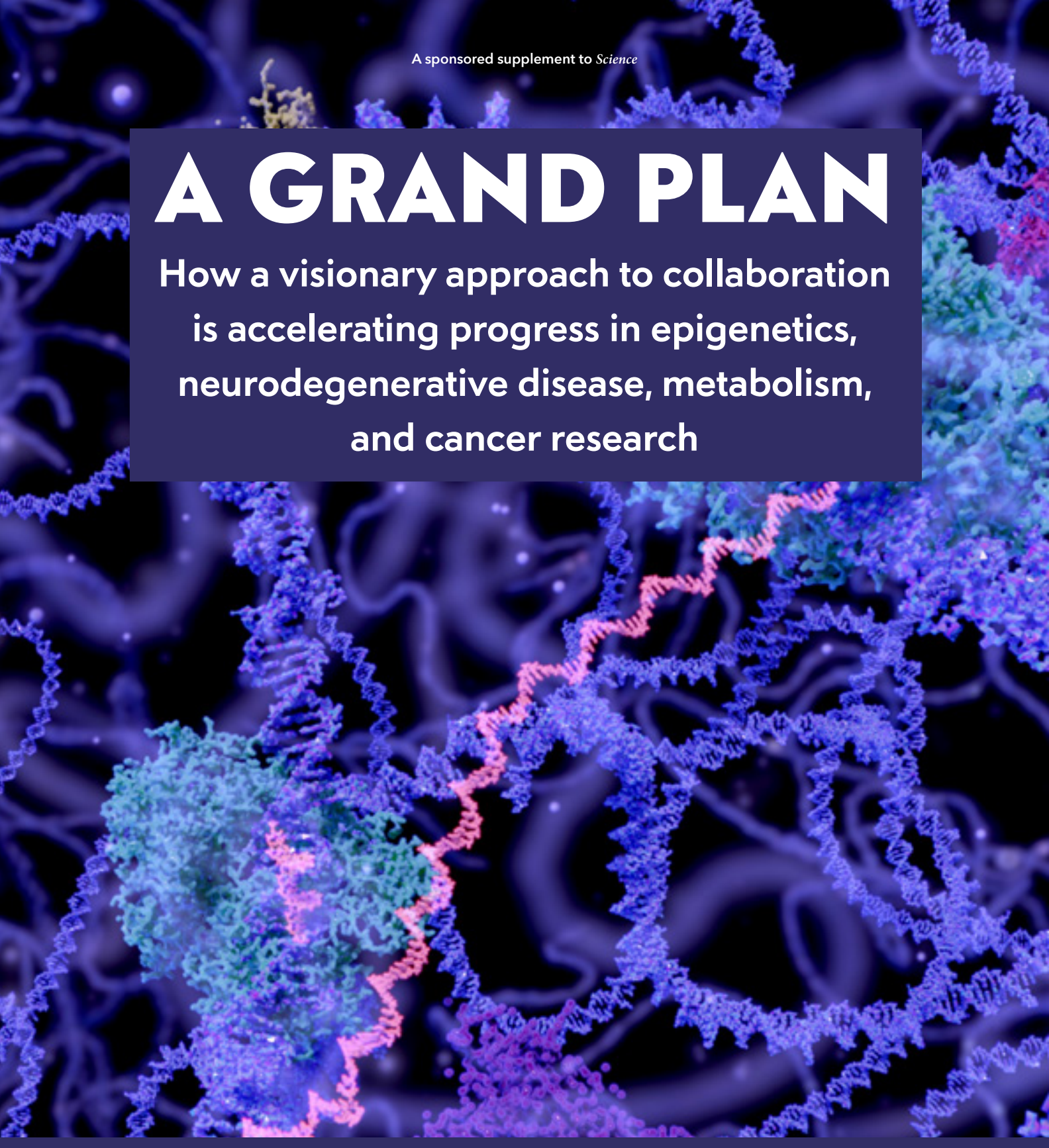
A sponsored supplement to *Science*

A GRAND PLAN

How a visionary approach to collaboration is accelerating progress in epigenetics, neurodegenerative disease, metabolism, and cancer research



Sponsored by

VAN ANDEL INSTITUTE®

Produced by the *Science*/AAAS Custom Publishing Office

Science | AAAS

Table of Contents

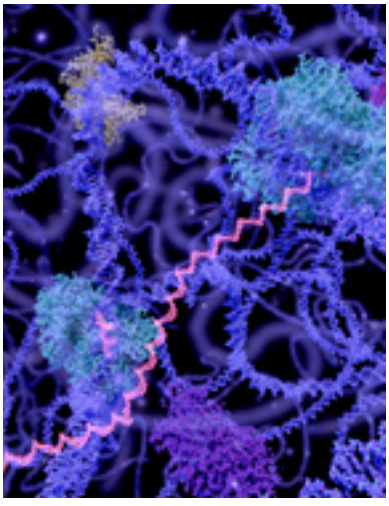


IMAGE: © JUAN GAERTNER/SHUTTERSTOCK.COM

This supplement was produced by the *Science*/AAAS Custom Publishing Office and sponsored by Van Andel Institute.

The original content articles that appear in this supplement have not been peer-reviewed nor have they been assessed by *Science*. Articles can be cited using the following format: [AUTHOR NAME(S)] [CHAPTER TITLE] in A Grand Plan: How a visionary approach to collaboration is accelerating progress in epigenetics, neurodegenerative disease, metabolism, and cancer research (*Science*/AAAS, Washington, DC, 2023), p. [xx-xx].

Editor: Erika Gebel Berg, Ph.D.
Proofreader/Copyeditor: Morgan Everett
Designer: JD Huntsinger

Chris Hoag
Sales Manager
choag@aaas.org
+1 (202) 590-8176

15 December 2023

Different is good

For all the ingenuity that scientists exhibit, the way science gets done in many traditional institutions can be a bit stuffy, slow, and bureaucratic. That's why it's been so refreshing and inspiring to get to know Van Andel Institute (VAI), a place that just does science differently. From my own days in the lab, I remember how stressful just dealing with administrative tasks could be and how long it took to get anything substantial done. VAI has efficiency built into its DNA to make it as easy as possible for its scientists to do science. This doesn't just make the scientists' lives easier, it can bring about scientific advances faster, and for people with life-threatening illnesses, the stakes are high.

VAI is tackling some of the most fearsome and challenging diseases humans can face in their lifetime: cancer, neurodegeneration, and metabolic diseases. And while these diseases may seem like they have nothing to do with each other, VAI is fostering breakthroughs by looking at what these diseases share and how advances in one condition can bring about progress in another. One central theme: epigenetics. This field explores how the environment and people's behaviors impact how genes work. Epigenetic changes play a role in many diseases. The big idea at VAI is that by creating a place that elevates epigenetics and its intersection with different diseases, scientists have a better chance of accelerating progress toward developing treatments with broad reach.

And there is no better place to study epigenetics than at VAI. In this booklet, you'll meet epigenetics pioneer and VAI's Chief Scientific Officer Peter Jones, who recruited a world-class team of scientists that turned VAI into an epigenetic epicenter. But you'll also hear from scientists from across VAI about their research and their experience in Grand Rapids. Collaboration is woven into the fabric of how VAI gets work done, offering ample opportunities for scientists to learn from each other and join together toward a common goal. We've included three articles from the *Science* family of journals, work that epitomizes the collaborative spirit at VAI as well as the cross-disciplinary approaches that are driving research forward. And you really can't talk about VAI without talking about Grand Rapids, Michigan. Make sure to check out the article on the last page of this booklet to learn about this growing city with a big personality, but be warned: you may find yourself packing your bags and starting a new life in a grand place that does serious science.

Erika Gebel Berg, Ph.D.
Director and Senior Editor
Science/AAAS custom publishing

PHOTO: VAN ANDEL INSTITUTE

Collaboration and innovation are the catalysts for discovery

At the heart of science is the drive to discover, to more clearly understand our world and ourselves. The resulting breakthroughs are the bricks that help us build a better future, one in which "eureka!" moments in the lab are translated into life-changing prevention, diagnostic, and treatment strategies.

Improving health for generations to come has been Van Andel Institute's mission since it was founded by the Van Andel family more than 25 years ago in Grand Rapids, Michigan. Our strategy is simple: provide the support and infrastructure required for our scientists and collaborators to ask bold questions, conduct rigorous research, and unleash innovation.

The fundamental mechanisms that drive health and disease are immensely complex and require us to combine expertise from multiple fields to make substantive advances. Science cannot—and should not—be siloed. Instead, we must work together to ask the big questions, challenge our assumptions, and explore critical problems from every angle.

We are at a transformative new frontier in science. New technologies, and the immense data sets they generate, are illuminating insights at a rapid pace. At the same time, discoveries across the basic, translational, and clinical realms are refining—and sometimes upending—longstanding knowledge. We must continue to work together across disciplines to realize the promise that these opportunities provide: a healthier future for all.

In this supplement, you'll learn more about our team-focused approach and the collaborative, impactful research it enables across our areas of focus: epigenetics, neurodegeneration, metabolism, cell biology, and structural biology. We also share the blueprint for what makes Van Andel Institute uniquely suited to supporting truly transformational research.

We are growing, both in size and impact. In the coming years, our upward trajectory is bound to accelerate as we recruit new talent and enhance our areas of strength, all with an eye toward translating promising discoveries into new therapies. We hope that sharing our story will reinforce the importance of collaboration and shine a spotlight on interdisciplinary innovation. Through this supplement, we invite you to learn more about us and join our collective mission toward a better tomorrow.

Peter A. Jones, Ph.D., D.Sc. (hon)
Chief Scientific Officer
Van Andel Institute

Q&A with Van Andel Institute CSO Peter Jones

Can you say a bit about what first brought you to Van Andel Institute?
Joining Van Andel Institute offered a rare opportunity to build an epigenetics-focused research program. The Institute was doing great work and had strong foundations—I arrived with the goal of taking all of that to the next level.

What impressed you about VAI?
One of the great things about the Institute that we've continued to build over the years is the close relationship between research and operations. Our Ops team is second to none and ensures our scientists have everything they need to succeed. Dr. Jana Hall, the chief operating officer—who, by the way, has a Ph.D. in human genetics—really understands the function of operations as a means to support great science.

The Institute also is fortunate to have a wonderful building and an exceptional philanthropic foundation from the founding family, the Van Andels. Their commitment ensures the Institute has the resources needed to be nimble and leading-edge.

What are your priorities as CSO?
I came to VAI with the express goal to develop the best epigenetics program in the world. Today I'm confident saying that we are among the best. We started with the premise that we wanted to recruit more senior investigators to help build the program—and that strategy proved to be immensely successful. And then, of course, we recruited exceptional, rising-star junior faculty to further build the program. Epigenetics is a big field—if you really want to understand how it works, you need broad representation of disciplines to see how it all fits together.

What do you think makes Van Andel Institute appealing from a recruiting perspective?
We have a strong infrastructure that supports great science. The alignment between research and operations means that we can shift the administrative burden off our faculty while still ensuring everything is done correctly and to the letter. We also pay investigator salaries and provide each lab with annual funding, which enables more focus on research. Our investigators still are expected to obtain external grant funding, which validates the work we do and allows us to direct internal funds to support broader initiatives and lab funding.

And what about resources like core facilities?
More than a third of experiments at institutes like ours are done in collaboration with core facilities. Far less experimentation is done at your own bench, where the focus is often on producing samples that are then analyzed by core facilities. Many of those are foundational things like the vivarium, for example, or the genomics facility. Optical imaging is very strong. But we're also an early adopter of cryo-electron microscopy—we set up a suite here about 6 years ago that gives us atomic-scale imaging capabilities. We have a world-class Pathology and Biorepository Core. Many of these big national activities like the Cancer Moonshot or the Cancer Proteomic Tumor Analysis Consortium program require a consolidated collection of biospecimens, and we are a contractor to those activities. Our staff has deep expertise in developing, managing, and processing sample kits to ensure samples are collected and stored under appropriate conditions.

How has Van Andel Institute been extending opportunities for aspiring and early-career scientists?
We deeply value our early career scientists and our actions reflect that. A positive postdoctoral fellowship goes far beyond research experience. Dr. Erica Gobrogge established and runs our Office of Postdoctoral Affairs. Her full-time job is to support postdocs through their fellowships and help us recruit postdocs to the Institute. We also recognize the challenges that impact early career scientists and are committed to mitigating these issues. For example, we were among the first institutes in the U.S. to raise postdoctoral starting salaries to $70,000 a year.

We also have Van Andel Institute Graduate School, which leverages a training approach called problem-based learning. We're excited to welcome a new dean, Dr. Eric Swindell, who was associate dean at Baylor College of Medicine, Houston. He's an outstanding addition to our faculty and will continue elevating the graduate school.

What is your take on how VAI has grown during your time there?
VAI has essentially doubled in size since I arrived nearly a decade ago. It's quite remarkable. I think we're an institute with extremely strong basic science that has the resources—intellectual, physical, and core—and the will to work with people all over the world to answer important questions that impact human health. That willingness to think big, to think globally, has made a huge difference. I can't wait to see what the next 10 years bring.

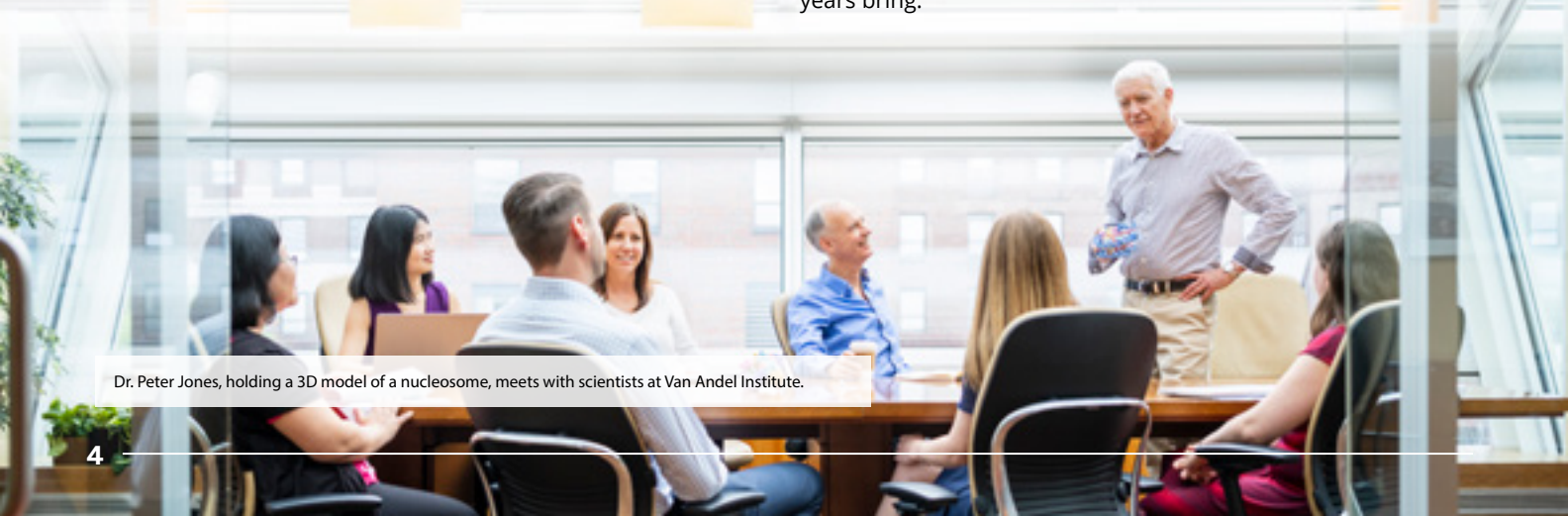
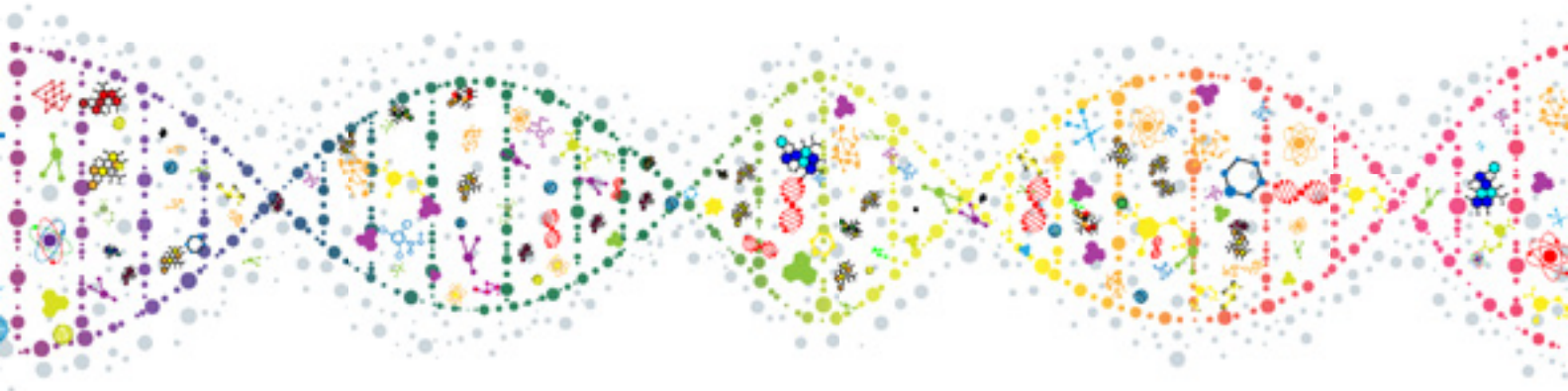


PHOTO: VAN ANDEL INSTITUTE

Dr. Peter Jones, holding a 3D model of a nucleosome, meets with scientists at Van Andel Institute.

Starting small, thinking big

How strategy and collaboration made Van Andel Institute an epigenetics epicenter



When epigenetics pioneer Peter Jones initially visited Van Andel Institute in 2013, one of the first things he noticed was the striking blown-glass DNA helix adorning its lobby. Tucked in amidst the blue orbs that comprised the helix's structure were a few red ones. They reminded Jones of the way methyl groups—epigenetic marks that annotate and regulate DNA—often are depicted in figures.

"It was a sign of sorts," said Jones, who went on to join VAI as chief scientific officer soon after. "As I learned more about the Institute, I realized that there was an incredible opportunity to create something truly transformative here in the epigenetics space."

Now, more than a decade later, Jones and his colleagues have built VAI into an epigenetics powerhouse, with complementary strengths in neurodegenerative science, metabolism, cell biology, and structural biology. Faculty routinely publish leading-edge epigenetics breakthroughs and contribute expertise to large-scale, federally led projects. The Institute is home to one of only two National Cancer Institute T32 grants that support post-doctoral training in cancer epigenetics and the only thematic SPORE grant focused on cancer epigenetics, a project co-led by Coriell Institute for Medical Research. On the translational front, VAI supports a consortium that brings together scientists, physicians, and organizations to move promising epigenetic combination therapies for cancer into clinical trials.

VAI's rapid ascent is no accident. It is the result of strategic vision, a commitment to catalytic collaboration, and a laser focus on ensuring its scientists and collaborators have the support and infrastructure needed to succeed. This ambitious drive has been present since the Institute's founding in 1996, with Chairman and CEO David Van Andel building on his parents' vision for a wellspring of innovation in Grand Rapids, Michigan that improves health for people around the globe. Its robust endowment, collaborative atmosphere, and nimble organizational structure have enabled the Institute to punch well above its weight in terms of scientific output. This progress is spurred on by VAI's ability to implement state-of-the-art technologies and draw top-notch scientific talent.

Jones is proud of how much the Institute has accomplished in a relatively short time. Since he joined in 2014, VAI has nearly doubled in size and grown in reputation. "I think we've hit a critical mass, and we're ready to go to the next level," he says. Compared to many larger academic operations, VAI remains a relatively small research institute—with more than 500 scientists, educators, and support staff. Its size is part of a visionary institutional strategy that emphasizes cooperation, funding support, and operational efficiencies to fuel exciting interdisciplinary collaborations at VAI and beyond.

An eye on epigenetics

Building an epigenetics program from scratch was no small feat. One of Jones' first steps after arriving at VAI was to assemble an all-star team of scientists focused on understanding how diverse biological processes interact with epigenetics, with the goal of informing novel therapies that save lives. The Institute has continued to recruit heavily and now the epigenetics faculty comprise its largest department.

As a foundational "on-off" switch for the genome, epigenetic processes are coupled to virtually every system in the body. Although epigenetics is studied in organizations around the world, few have committed resources across the epigenetics research continuum—from basic to translational and clinical—like VAI has. This full-spectrum approach has resulted in a gravitational pull that only continues to increase. For example, Yvonne Fondufe-Mittendorf joined VAI due in large part to its critical mass of scientists in the field. "I was the only epigeneticist at my previous institution," she says. "Here at Van Andel Institute, every day I talk to people about epigenetics and hear new things that could enhance my science."

Fondufe-Mittendorf's work principally focuses on understanding how environmental factors, particularly toxic heavy metals like arsenic, impact cellular function via epigenetic mechanisms. Arsenic is already well known for its ability to cause damaging and potentially carcinogenic mutations in the genome. "But at the levels that most people are exposed to, it doesn't really cause a mutation or causes very low levels of mutations," says Fondufe-Mittendorf. Arsenic remains harmful and toxic even

at these levels, and her group has shown that cells undergoing arsenic-induced carcinogenesis exhibit striking changes in their pattern of DNA methylation—one of the most important subtypes of epigenetic modification. Some of her current work is exploring how arsenic directly influences the binding of a protein called CTCF that ultimately governs methylation patterns in the genome, which could give this toxicant far-reaching effects on global gene expression.

Epigenetics research at VAI is not confined to a single department. Rather, it is a common thread that winds through all the Institute's research areas. Metabolic processes, for example, directly influence epigenetics in a number of ways, including providing the raw building blocks that are subsequently used to enzymatically mark DNA and histones. "All of the 'ink' to lay down those marks comes from metabolism," says Russell Jones, chair of VAI's Department of Metabolism and Nutritional Programming. Russell Jones collaborates closely with his VAI colleague Connie Krawczyk to investigate how epigenetic factors inform the interplay between diet, metabolism, and immune function. In one recent study, Russell Jones, Krawczyk, and collaborators showed that the accumulation of molecules known as ketone bodies, which are a metabolic byproduct of fasting, promotes epigenetic changes that boost the function of pathogen- and tumor cell-killing T cells. Russell Jones believes this could be an evolutionary mechanism that helped protect our ancestors during times of food scarcity, but also sees clear advantages for fighting infections in the present day.

Another area ripe for exploration is the relationship between epigenetics and genetics—and detangling their effects from one another. This is especially important in diseases that develop over long periods of time, such as neurodegenerative disorders like Parkinson's and dementia. The etiology of these diseases is frequently rooted in a complex mix of epigenetic, genetic, and environmental factors. There are exceptions, however. Five to 10% of Parkinson's cases have a wholly genetic origin and understanding how these cases arise versus sporadic Parkinson's is crucial to sparking potentially life-changing advances, says Darren Moore, chair of VAI's Department of Neurodegenerative Science.

VAI's research areas are far from siloed—instead, the Institute encourages investigators to think outside the box and leverage interdisciplinary collaborations to tackle pressing research questions.

For example, Fondufe-Mittendorf has been in discussion with her colleagues working on neurodegenerative diseases at VAI to pursue a study investigating the epigenetic factors that contribute to the neurological damage seen in people with Gulf War Syndrome. "I wouldn't have done that in any other place, but because everyone here is so close, unexpected collaborations arise," she says. The still-enigmatic syndrome, identified among veterans of the 1990–1991 war in Iraq and Kuwait, is associated with a spectrum of debilitating symptoms including pain, fatigue, and deficits in memory and cognition. While trauma is thought to play a role in Gulf War Syndrome, it may not be the only factor. "Now we are seeing that they were exposed to things that they burned, and all of those contain heavy metals," she says.

Similarly, when an unexpected finding pops up in a lab's routine research, the right partner for a collaboration could be directly down the hall. Krawczyk's lab has been studying a protein called KDM5C, which selectively strips methyl modifications off chromosomal proteins. To aid in their work, her lab generated a mouse strain that lacks the gene encoding KDM5C. "My postdoc, Lukai, was very observant and he noticed the bones of these mice looked kind of weird," she says. This prompted her group to team up with the lab of Tao Yang, a colleague in VAI's Department of Cell Biology who specializes in skeletal development. Their collaboration revealed that KDM5C introduces epigenetic changes that stimulate the activity of bone-degrading cells known as osteoclasts. KDM5C is expressed by a gene on the X chromosome, and thus exerts a stronger effect in women. Krawczyk and Yang are looking into opportunities to avert osteoporosis in older women with treatments that modulate KDM5C.

Another example of collaborative connection may be found in a 2020 paper from Peter Jones and colleagues in the Institute's Department of Structural Biology, which visualized the structure of nucleosome-bound DNMT3A and DNMT3B, both critical DNA methyltransferases that are often dysregulated in cancer. The project was a serendipitous mix of VAI's expertise in both epigenetics and structural biology coupled with its

Dr. Yvonne Fondufe-Mittendorf

PHOTO: VAN ANDEL INSTITUTE

leading-edge technology, namely its state-of-the art cryo-electron microscopy (cryo-EM) suite. Additional studies are underway, thanks in part to Jones' recently renewed NCI Outstanding Investigator Award, which includes cryo-EM as a key method for parsing the relationship between epigenetic markers like DNMT3A and cancer.

All these projects share another common thread: collaboration. Teamwork is at the heart of VAI's strategy, and the Institute fosters it both inside and outside of its walls.

Catalyzing clinical collaboration

At their most impactful, collaborations span laboratories, fields, technologies, and organizations, combining strengths and expertise from across disciplines. Perhaps the most mature example of VAI's commitment to fostering extensive collaborations, according to Peter Jones, is the Institute's clinical trial programs to accelerate the evaluation of promising new therapies in cancer and Parkinson's disease.

Since 2014, the Van Andel Institute–Stand Up To Cancer Epigenetics Dream Team has brought together scientists, physicians, medical centers, and companies to evaluate potential cancer treatments that combine epigenetic drugs with other agents. Establishing the team, which was the first of its kind to be self-funded, was a key factor in Jones' decision to come to VAI, in part because it provided a way to translate discoveries in epigenetics into potential cancer treatments.

One recently concluded phase 1/2 trial supported by the team demonstrated that pairing an epigenetic drug known as a hypomethylating agent with an immune system-activating drug called atezolizumab extended the survival of people with relapsed or refractory myelodysplastic syndrome (MDS) by an average of 10 months compared to historical controls. Patients receiving this combination experienced a median overall survival of 15 months, compared to five months for the historical control cohort.

VAI also collaborates with the UK research charity Cure Parkinson's on the International Linked Clinical Trials Program, which pursues disease-modifying treatments for Parkinson's by exploring drugs developed for other conditions. To date, the consortium has supported a host of trials, including a phase 3 trial of the diabetes drug exenatide and the impending phase 3 trial of the respiratory drug ambroxol.

"One of the strengths of International Linked Clinical Trials is the collaboration," says Moore, who is a member of the iLCT Scientific Committee. "It brings together scientists, physicians, and people with Parkinson's toward one unified goal—find a treatment that slows or stops Parkinson's progression. It's an urgent mission and collaboration is our best chance of getting there."

Smaller, better, faster, stronger

Collaboration and rigorous research require a supportive infrastructure that enables innovation. At VAI, that research-focused architecture is the domain of its expansive operations team.

"We have a much higher ratio of operations to research than a big academic institution," says Moore, and this means that labs can quickly get personalized support for preparing grant applications, setting up research protocols, and taking care of

PHOTO: VAN ANDEL INSTITUTE

Dr. Russell Jones (right) with graduate students Brandon Oswald (center) and Shelby Compton (left)

other processes that could take valuable time away from the lab. Critically, this agility does not mean cutting corners. "The scientists work closely with operations and compliance staff, which enables scientists to spend more time focusing on science," says Peter Jones, citing in particular the Institute's robust research protections infrastructure.

This operational efficiency also translates to greater speed in building out the technological capabilities that makes modern research possible. VAI is currently home to nine core facilities, which provide easy access to instrumentation that would be far too expensive, complicated, or labor-intensive for any individual lab to operate on its own.

But it isn't just the access to technology and staff expertise that's important. It's also the Institute's flexibility and responsiveness when new needs arise.

When he joined VAI, Russell Jones was deeply excited about the prospect of building out a dedicated mass spectrometry facility, but he was also concerned about managing his expectations. Such a facility would be a boon not only for his personal work, but such versatile instrumentation could elevate science across the Institute. Mass spectrometers are powerful tools for profiling the biochemical fingerprints of human physiology and metabolism—a central component of Russell Jones's research—but they require lots of specialized infrastructure, and this takes time. At another academic institution, Russell Jones recalls, "I waited a year and a half to have a wall removed to put in a new mass spec." Accordingly, he was cautiously optimistic that VAI would get things up and running within 5 years.

Instead, it took just two, a remarkable feat of institutional agility and efficiency. VAI's cryo-EM suite followed a similar storyline: faculty identified a need and VAI leadership and staff moved quickly to develop a sound strategy to make it happen.

"Ideas for new cores, for example, have started in faculty meetings," Moore says. This was how the mass spectrometry facility plan got off the ground so quickly. And in addition to providing a valuable resource for VAI researchers, Russell Jones notes that this facility has been a big draw for talented faculty in the metabolism research space, including recent hires Evan Lien and Sara Nowinski.

The next frontier

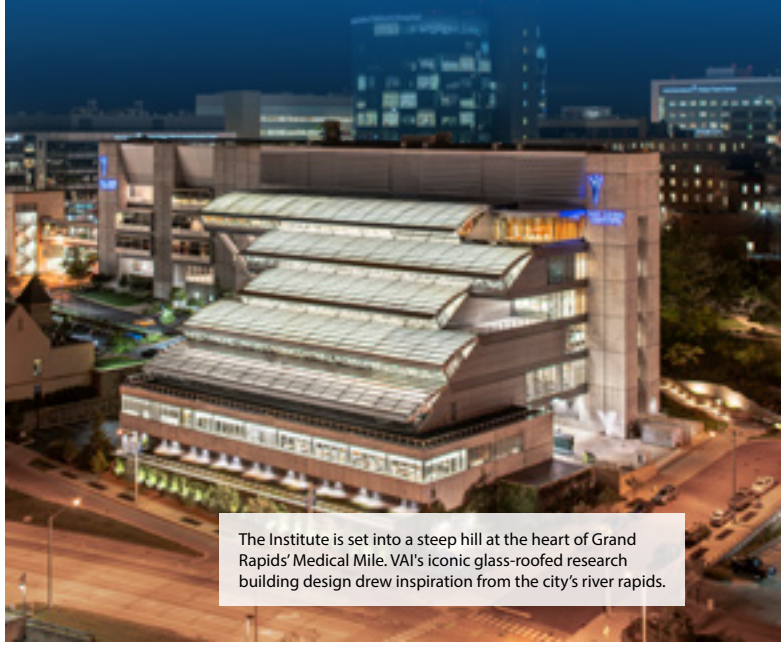
This rapid growth in expertise, technical capabilities, and research impact has not escaped the scientific community's notice. "At metabolism conferences, people's eyes light up because we have the most spectacular mass spec core you could imagine," says Krawczyk. Likewise, Peter Jones recalled a recent conference in which VAI's epigenetics faculty comprised a substantial portion of the speaker roster.

While staying true to the values that made it a world-class institution, VAI continues to grow, adapt, and collaborate. This will inevitably include building on the Institute's core strengths in epigenetics, neuroscience, metabolism, structural biology, and cell biology. But otherwise, CSO Peter Jones is keeping an open mind regarding what comes next and will only commit to tackling future research endeavors with the same agility, creativity, and focus that have already brought VAI this far. "We're all aligned behind our mission to improve human health—and everything we do reflects that," he says. "That's what makes this institute special."

VAI at a glance

- VAI was founded by the Van Andel family in 1996 in Grand Rapids, Michigan.
- The Institute is home to more than 500 scientists, educators, and support staff.
- VAI ranks #27 in the Nature index Biological Sciences category for NGOs/nonprofits and #46 overall.
- Between 2014 and 2023, the number of publications from VAI researchers in high-impact journals (>7) rose more than 200%.
- VAI is among the top 5 highest funded research organizations in Michigan.
- As of September 2023, VAI has supported the launch of more than 40 clinical trials in cancer and Parkinson's disease.

PHOTO: VAN ANDEL INSTITUTE



The Institute is set into a steep hill at the heart of Grand Rapids' Medical Mile. VAI's iconic glass-roofed research building design drew inspiration from the city's river rapids.

NEUROSCIENCE

Computational modeling of tau pathology spread reveals patterns of regional vulnerability and the impact of a genetic risk factor

Eli J. Cornblath[1,2], Howard L. Li[3], Lakshmi Changolkar[3], Bin Zhang[3], Hannah J. Brown[3], Ronald J. Gathagan[3], Modupe F. Olufemi[3], John Q. Trojanowski[3], Danielle S. Bassett[1,4,5,6,7,8], Virginia M. Y. Lee[3], Michael X. Henderson[9]*

Neuropathological staging studies have suggested that tau pathology spreads through the brain in Alzheimer's disease (AD) and other tauopathies, but it is unclear how neuroanatomical connections, spatial proximity, and regional vulnerability contribute. In this study, we seed tau pathology in the brains of nontransgenic mice with AD tau and quantify pathology development over 9 months in 134 brain regions. Network modeling of pathology progression shows that diffusion through the connectome is the best predictor of tau pathology patterns. Further, deviations from pure neuroanatomical spread are used to estimate regional vulnerability to tau pathology and identify related gene expression patterns. Last, we show that pathology spread is altered in mice harboring a mutation in leucine-rich repeat kinase 2. While tau pathology spread is still constrained by anatomical connectivity in these mice, it spreads preferentially in a retrograde direction. This study provides a framework for understanding neuropathological progression in tauopathies.

INTRODUCTION

Neurodegenerative diseases, including Alzheimer's disease (AD) and Parkinson's disease (PD), are estimated to affect over 60 million people worldwide (*1*, *2*). Neurological symptoms and the presence of pathological protein inclusions are used to categorize the two diseases. However, there exists substantial overlap in both symptoms and pathologies, especially as these diseases progress (*3*, *4*). Tau pathology, while diagnostic of AD and other primary tauopathies, appears prominently in PD, PD dementia (PDD), and dementia with Lewy bodies, where it correlates with α-synuclein pathological burden and cognitive decline (*5*, *6*). These data suggest that multiple pathologies may act additively to influence disease progression and that underlying risk factors for one pathology may confer risk for additional pathologies.

Postmortem neuropathology studies have demonstrated that patients with more severe clinical AD and PDD have elevated levels of pathological tau in an increasing number of regions of the brain (*5*–*8*). The stages of observed tau pathology (beginning in the locus coeruleus, then transentorhinal and entorhinal cortex, and moving through the hippocampus and cortical regions) are suggestive of pathology "spreading" (*7*, *8*). Pathological tau from human brains injected into nontransgenic (NTG) mice can be internalized into nearby neurons, initiating misfolding and hyperphosphorylation of endogenous mouse tau in a prion-like manner into intraneuronal inclusions resembling those from human disease tissue (*9*–*11*). Over time, tau pathology can be found in more regions of the mouse brain connected to the injection site (*9*–*11*), suggesting that tau is spreading. How tau pathology spreads has been a matter of debate because it has not been possible to disambiguate the contribution of neuroanatomical connectivity, spatial proximity of regions, and intrinsic neuronal vulnerability. Recent studies based on mathematical modeling of human brain suggest that anatomical connectivity serves as a strong predictor of brain atrophy or general pathology patterns in neurodegenerative diseases (*12*–*17*).

In the current study, we investigated processes underlying the development and spread of tau pathology. We developed a methodology for reproducibly quantifying tau pathology in 134 regions of the mouse brain following an intracranial injection of pathogenic tau. Tau pathology in this model begins slowly, first affecting the injection site and highly connected regions. As time progresses, more regions are affected. Once affected, regions show a nearly linear increase in tau pathology over time. While this atlas of pathology is informative, it is still not possible to discern mechanisms of pathology spread without considering the brain as a network of interconnected regions. We therefore used computational analysis of spatiotemporal tau pathology patterns. We found that tau pathology spread is best explained by diffusion along neuroanatomical connections in a bidirectional manner. Residual variance in the data that was not well explained by connectivity was used to generate estimates of differential vulnerability to tau in measured regions. Comparison of regional vulnerability to regional gene expression identified several previously unidentified candidate genes that may control susceptibility to tau pathology. To further assess the utility of quantitative pathology and network modeling, we analyzed a mouse expressing the LRRK2^{G2019S} mutation. This mutation is a risk factor for PD, but many of these patients exhibit tau pathology. LRRK2^{G2019S} mice exhibit a bias toward retrograde spread of tau pathology, providing

[1]Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA 19104, USA. [2]Department of Neuroscience, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [3]Institute on Aging and Center for Neurodegenerative Disease Research, Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [4]Department of Electrical and Systems Engineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA 19104, USA. [5]Department of Physics and Astronomy, College of Arts and Sciences, University of Pennsylvania, Philadelphia, PA 19104, USA. [6]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [7]Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [8]Santa Fe Institute, Santa Fe, NM 87501, USA. [9]Department of Neurodegenerative Science, Van Andel Institute, Grand Rapids, MI 49503, USA.
*Corresponding author. Email: michael.henderson@vai.org

insight into the network-level impact of cell biology events. This work provides a framework for understanding the spread of pathological tau throughout the brain and investigating the impact of genetic risk factors on pathology progression.

RESULTS

Quantitative immunohistochemistry to evaluate pathological tau spread

To investigate mechanisms underlying the spread of tau pathology, we used a recently developed seed-based model of tauopathy in which injection of AD brain–derived tau into NTG mice can induce the misfolding of endogenous tau into hyperphosphorylated tau inclusions in a time- and region-dependent manner (*9*). Previous studies have suggested that this spatiotemporal pattern of tau pathology induction is consistent with spread along neuroanatomical connections (*9*, *10*), although these patterns were not quantitatively evaluated. Biochemical sequential detergent extraction of gray matter from AD patient brains was used to obtain an enriched fraction of paired helical filament (PHF) AD tau (Fig. 1A) (*9*). This extraction method yielded a final purity of 22.6 to 35.7% tau, with 0.01% or less α-synuclein and amyloid β (table S1). The purified tau fraction retains the pathogenic conformation present in human disease and induces the misfolding of tau in mice without the overexpression of tau (*9*). To investigate how tau pathology spreads through the brain, we injected NTG mice with PHF tau from individual extractions in the hippocampus and overlaying cortex (Fig. 1B). Mice were allowed to age 1, 3, 6, or 9 months post-injection (MPI) to capture the temporal dynamics of tau pathology spread (Fig. 1C). Brains were then sectioned, and representative sections (Fig. 1D) were selected and stained for phosphorylated tau pathology throughout the brain. We manually annotated 194 areas from 134 anatomical regions on the selected sections (Fig. 1E and fig. S1) so that pathology could be quantified as the percentage of each area occupied by pathology (Fig. 1, F and G). Brain region annotations were based on the Allen Brain Atlas (ABA) Common Coordinate Framework (CCF v3, 2017, brain-map.org), although smaller subregions with minimal pathology were grouped to minimize error and annotation time. To mitigate overinterpretation of individual sections, we similarly annotated and quantified a second set of nearby sections (16,684 annotations in total), and the average for each region was used for subsequent analyses.

Quantitative pathology mapping reveals dynamic patterns of tau pathology spread over time

Because imaging of tau pathology in these mice can only be done after death, data are pseudo-longitudinal. That is, different time points are represented by separate groups of mice. Despite this fact, tau

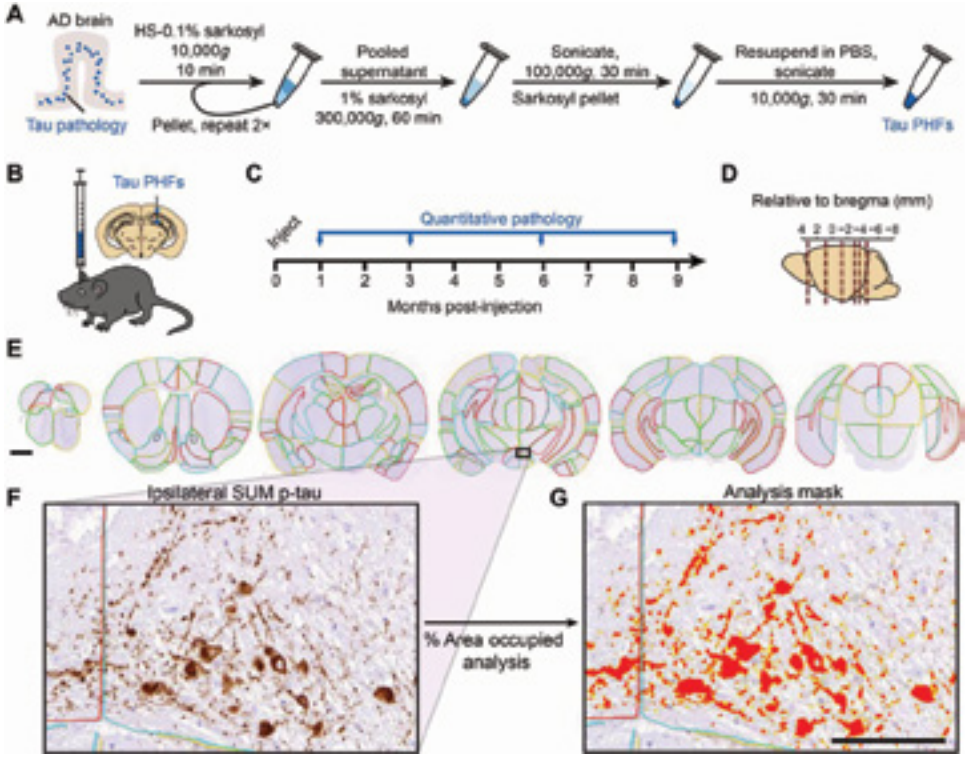


Fig. 1. Quantitative immunohistochemistry to evaluate pathological tau spread. (**A**) AD brain with a high burden of tau pathology was chosen for extraction of pathological tau. Brains went through sequential extraction of tau PHFs as noted in the schematic. Final tau PHF preparations were used for all subsequent steps. PBS, phosphate-buffered saline. HS, high salt. (**B**) NTG mice were injected unilaterally with AD PHF tau in the hippocampus and overlaying cortex as shown at 3 to 4 months of age. (**C**) Mice were euthanized 1 ($n = 4$), 3 ($n = 8$), 6 ($n = 6$), or 9 ($n = 6$) months following injection. (**D**) Mouse brain was sectioned, and the sections representing the regions shown were stained for pathological tau. (**E**) Representative sections were selected, and 194 regions were annotated for each brain. A second set of nearby sections was similarly annotated to reduce selection bias. Scale bar, 1 mm. Annotation colors are arbitrary. (**F**) An enlarged image of the annotated supramammillary nucleus (SUM) is shown with the inclusions stained for pS202/T205 tau. (**G**) Annotations allow automated quantification of percentage of area occupied with pathology in specific regions of the brain. An analysis mask is overlaid on the image in (F) to demonstrate this quantification of pathology. Scale bar, 100 μm.

pathology patterns were highly reproducible across cohorts and time points (Fig. 2). One remarkable finding from this analysis was that once pathology seeds a region, it continues in an almost linear fashion over time in most regions (Fig. 2), suggesting that the process of pathology formation is similar across the brain. The lack of a plateau in most regions also suggests that this model is recapitulating the early disease process, before the saturation of pathology and profound neuron death. While the process of pathology formation was similar across regions, the delay between injection and pathology formation varied widely. Regions at the injection sites or with high connectivity to the injection sites developed pathology by 3 MPI (Fig. 2, A and B). Other regions showed a further delay, with only minimal pathology before 6 MPI (Fig. 2, C and D). Last, some regions did not develop pathology until 9 MPI (Fig. 2, E and F), and some regions did not develop substantial pathology during the course of the study (fig. S10).

While patterns of pathology in individual regions are informative, overall patterns throughout the brain are best observed as a heatmap overlaid on the mouse brain (Fig. 3) or as a region-by-time matrix (fig. S2). Several whole-brain patterns were apparent. At 1 MPI, there was minimal tau pathology outside of the injection sites. At 3 MPI, more pathology accumulated ipsilateral to the injection site including hippocampal and entorhinal regions. By 6 MPI, pathology spread to the contralateral hippocampus and associated cortical areas. By 9 MPI, pathology continued to spread, affecting more rostral regions, although certain rostral cortical and contralateral thalamic

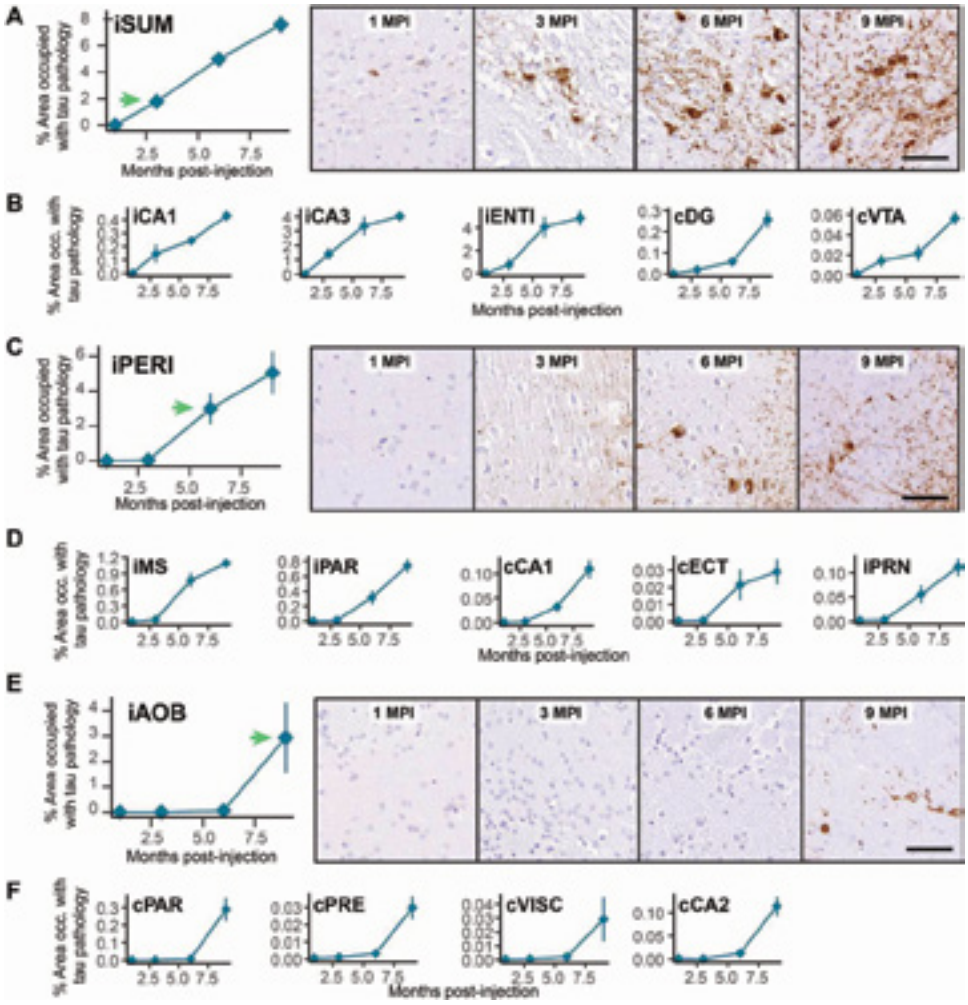


Fig. 2. Brain regions show progressive accumulation of tau pathology. The percentage of area occupied with tau pathology is plotted as a function of time demonstrating three distinct onsets of pathological tau accumulation: 3 MPI (A and B), 6 MPI (C and D), and 9 MPI (E and F). Most of the other regions also fall into one of these patterns. Green arrows denote the time point with initial substantial pathology accumulation. (**A**) The ipsilateral supramammillary nucleus accumulates pathology by 3 MPI as demonstrated by the quantitative pathology plot and images at right. (**B**) Additional regions that show a similar onset of pathology at 3 MPI (iCA1, field CA1 of the hippocampus; iCA3, field CA3 of the hippocampus; iENTl, entorhinal area, lateral; cDG, dentate gyrus; cVTA, ventral tegmental area). (**C**) The ipsilateral perirhinal area (iPERI) accumulates pathology by 6 MPI as demonstrated by the quantitative pathology plot and images at right. (**D**) Additional regions that show a similar onset of pathology at 6 MPI (iMS, medial septal nucleus; iPAR, parasubiculum; cCA1, field CA1 of the hippocampus; cECT, ectorhinal area; iPRN, pontine reticular nucleus). (**E**) The ipsilateral accessory olfactory bulb (iAOB) accumulates pathology by 9 MPI as demonstrated by the quantitative pathology plot and images at right. (**F**) Additional regions that show a similar onset of pathology at 9 MPI (cPAR, parasubiculum; cPRE, presubiculum; cVISC, visceral area; cCA2, field CA2 of the hippocampus). Data are represented as means ± SEM. Scale bars, 50 μm.

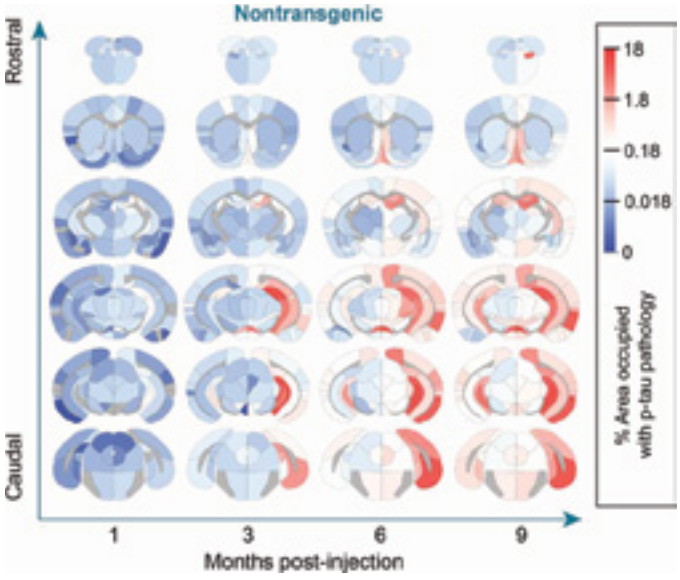


Fig. 3. Quantitative pathology mapping reveals dynamic patterns of tau pathology spread over time. Regional pathology measures plotted on anatomical maps as a heatmap, with blue representing minimal pathology, white representing moderate pathology, and red representing substantial pathology. Note that tau pathology was not quantified in white matter regions, and these regions are gray. Representative images for all regions can be found in the Supplementary Materials. *n* = NTG-1 M: 4, NTG-3 M: 8, NTG-6 M: 6, and NTG-9 M: 6.

regions show minimal pathology. Representative images of all regions are available in the Supplementary Materials (fig. S10). Together, these patterns are highly suggestive of tau pathology spread throughout the brain, which may involve transmission along neuroanatomical connections or other regional factors.

Tau pathology spread is predicted by diffusion through the neuroanatomical connectome

There is substantial evidence that tau can be released from neurons (*18*–*20*) in an activity-dependent manner (*21*, *22*) and internalized by other neurons (*23*–*26*), likely leading to the spread of tau pathology throughout the brain. However, it is still unclear precisely how spread occurs. The main hypotheses are that tau pathology spreads between anatomically connected regions of the brain, between physically contiguous regions, or to selectively vulnerable populations of neurons. Of course, these mechanisms of spread are not mutually exclusive; thus, we sought to assess how each of these putative routes contributes to the observed pathology spread in mice.

We evaluated the ability of a network diffusion model with neuroanatomical connectivity (brain-map.org) (*27*) as a scaffold to predict the empirical measures of tau pathology over time. ABA regions at the injection sites (iDG, iCA1, iCA3, iRSPagl, and iVISam) were used as seed regions for initiation of the model. Direct connectivity of these regions was highest in hippocampal, septal, and entorhinal regions of the brain (fig. S3A), and direct connectivity to the injection site was highly correlated with tau pathology measures in those regions (fig. S3, B and C). Our model posits that tau spreads from the injection site along anatomical connections both in retrograde and anterograde directions, with the final amount of regional pathology determined by a weighted sum of these two independent processes. This bidirectional anatomical connectivity model weakly predicted tau pathology at 1 MPI, likely due to minimal spread at this time point, and strongly predicted tau pathology at 3, 6, and 9 MPI (Fig. 4A). In addition, this bidirectional anatomical connectivity model outperformed a model in which tau spread was proportional to the Euclidean distance between regions (Fig. 4B). In the Euclidean distance plots, the five outlying regions in the upper right were the injection sites, as noted (Fig. 4B); however, inclusion or exclusion of these sites during model optimization did not measurably affect the fit of the Euclidean distance model (fig. S4, A and B). Excluding the injection sites from the bidirectional model did not impair its predictivity over the Euclidean distance model (fig. S4C). Rewiring the network to disrupt the connectome but preserve basic network properties such as in-degree (fig. S5A) or out-degree connectivity (fig. S5B) also reduced model performance. To further validate the anatomical connectivity model, we ensured that the model's performance was specific to the choice of the experimental injection site, compared with 500 randomly chosen sets of five regions with mean spatial proximity similar to that of the experimental injection sites (Fig. 4C). Models using the experimental seed regions were among the best performing models at all time points (Fig. 4C), confirming the specificity of our model to the experimental injection site. The performance of models using random sites could be partially explained by a generalized additive model relying on three variables: in-projection similarity between actual and alternate seeds, out-projection similarity between actual and alternate seeds, and Euclidean distance between actual injection sites and alternate seed sites (fig. S5C).

Last, we rigorously evaluated the out-of-sample performance of the bidirectional anatomical connectivity model compared to three other models in which spread was based on either Euclidean distance, anterograde spread alone, or retrograde spread alone. To compare the distributions of out-of-sample fits between each of the four models, we generated 500 train-test splits of the mice used in the study. Next, we obtained model parameters (diffusion rate constants and regression weights for anterograde and retrograde spread) in the training set and evaluated model performance in the test set at each time point. Our measure of performance was the spatial Pearson correlation coefficient between the observed pathology in the test set and the predicted pathology estimated by each model. This analysis revealed that the bidirectional model was superior to both the anterograde and retrograde models and that all three connectivity-based models were superior to the Euclidean distance model by 6 and 9 MPI (Fig. 4D). To understand whether the inclusion of Euclidean distance could improve predictivity of the bidirectional model, we evaluated a combined model that integrated bidirectional connectivity and Euclidian distance. We found that the bidirectional-alone model performed at least as well as the combined model at all time points (fig. S5D). To understand the contribution of anterograde and retrograde connections to the bidirectional model, we estimated the diffusion rate constant (fig. S5F) and standardized β (fig. S5G) from the model. We found a stronger contribution of retrograde connectivity to the spread model in both measures. In summary, these findings provide strong support for the notion that both anterograde and retrograde spread along anatomical connections independently contribute to the propagation of tau pathology, with retrograde spread being predominant.

Model residuals as an estimate of regional vulnerability

A third possible mediator of tau pathology spread is intrinsic neuronal vulnerability. While this concept is less easily parsed for

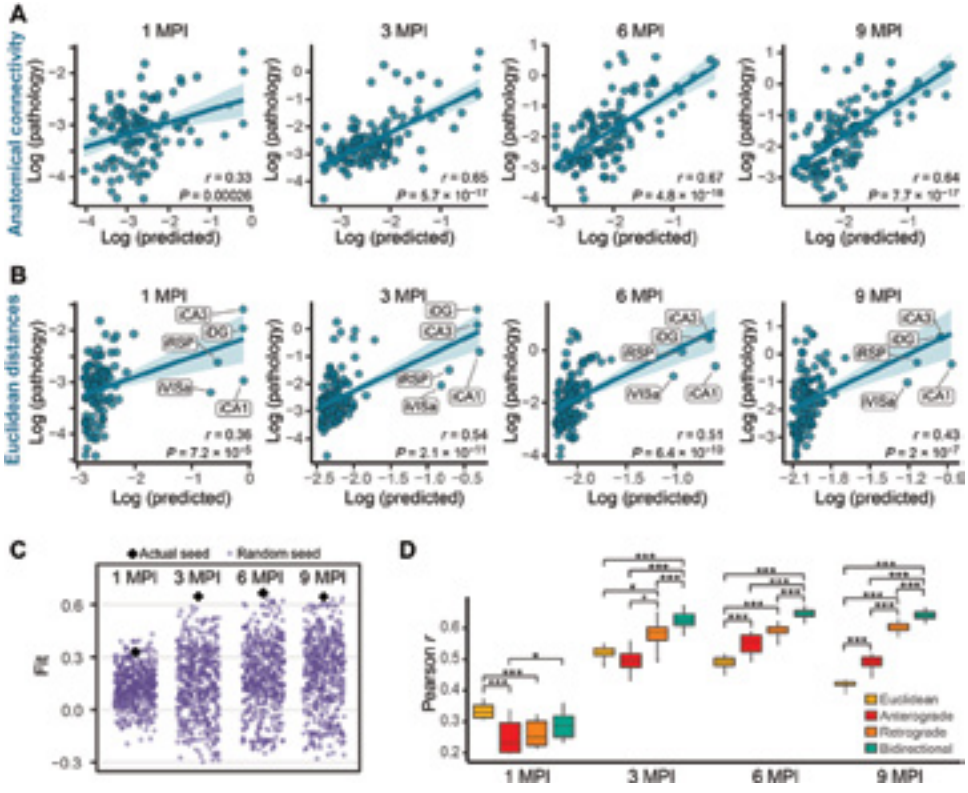


Fig. 4. Tau pathology spread is predicted by diffusion through neuroanatomical connectivity. (**A** and **B**) Predictions of log tau pathology (*x* axis) from spread models based on (A) retrograde and anterograde anatomical connections or (B) Euclidean distance, plotted against log actual regional tau pathology values (*y* axis) at 1, 3, 6, and 9 MPI. The solid lines represent the lines of best fit, and the shaded ribbons represent the 95% prediction intervals. The *r* and *P* values for the Pearson correlation between model-fitted values and observed pathology are noted on the plots. In (B), the injection site regions are noted but did not affect model fit (fig. S4). (**C**) To evaluate the specificity of the seed sites to predict the pathology spread pattern, 500 alternate combinations of five seed sites (purple dots) were evaluated for their ability to predict pathology spread at 1, 3, 6, and 9 MPI. Using the actual five injection regions (black diamonds) produced among the best fits at all time points (1 MPI, $P = 0.042$; 3 MPI, $P < 0.002$; 6 MPI, $P < 0.002$; and 9 MPI, $P < 0.002$. (**D**) Distributions of model fits in 500 held-out samples using Euclidean, anterograde, retrograde, and bidirectional models. Fit differences were analyzed by pairwise two-tailed nonparametric tests for different models. All connectivity models outperform Euclidean distance after 3 MPI, and a bidirectional model outperforms either a retrograde or anterograde model alone (*$P < 0.05$ and ***$P < 0.002$). In box plots, box edges represent the 25th and 75th percentiles, the middle line shows the median, and whiskers extend from the box edges to the most extreme data point value that is at most 1.5× interquartile range (IQR). Data beyond the end of the whiskers are plotted individually as dots.

tauopathies than for α-synucleinopathies where dopaminergic neurons appear particularly vulnerable, our network model of tau pathology spread can be used to infer regional vulnerability. We have previously validated this approach by showing that residuals from network model predictions of α-synuclein pathology correlated with α-synuclein gene expression, which was expected to contribute to regional vulnerability (*28*). Here, we estimated regional vulnerability to tau pathology spread by taking the residuals from the bidirectional anatomical connectivity model and averaging them over 3, 6, and 9 MPI (Fig. 5A). We excluded 1 MPI because of the low model predictivity and dissimilarity to residuals from other time points (fig. S6A). This estimate of regional vulnerability revealed that amygdalar, thalamic, and rostral cortical nuclei were resilient, while septal, mesencephalic, and caudal cortical regions were more vulnerable (Fig. 5A). We then tested whether tau gene (*Mapt*) expression from ABA in situ hybridization data was associated with vulnerability. *Mapt* is expressed quite broadly in the brain, including most gray matter and white matter regions, and showed no association with regional vulnerability estimates (fig. S6B), suggesting that tau expression is not a major limiting factor in the spread of tau pathology.

We sought to further investigate gene expression patterns associated with regional vulnerability by performing a genome-wide search for genes whose expression patterns measured by the ABA were spatially similar to our regional vulnerability measure. Many gene expression patterns correlated with regional vulnerability. After adjusting false discovery rate (FDR) to $q < 0.05$ (*29*), we found that 20 genes had expression patterns with a statistically significant spatial correlation with relative regional vulnerability to tau pathology (Fig. 5B). Twelve of these genes showed expression patterns that were negatively associated with regional vulnerability, and 8 of these genes showed expression patterns that were positively associated with vulnerability. Plotting the gene expression versus vulnerability of individual genes (Fig. 5, C and D, and fig. S6C) demonstrates that this association is not driven by outliers. In the future, gene expression patterns may be useful parameters of regional vulnerability in computational models of pathology spread.

In silico seeding from alternate sites

In addition to inferring mechanisms of spread through network modeling, we can also extend the value of our validated network

models by generating predictions of tau spread patterns from alternate injection sites, assuming that the rates of spread and contributions of anterograde versus retrograde spread are the same for those injection sites. For example, one of the earliest sites with tau pathology outside of the brainstem in humans is the entorhinal cortex (*7*). However, this site is difficult to inject reproducibly in mice because of its lateral location. In silico modeling of tau pathology spread from this injection site shows a more lateralized spreading pattern that largely affects hippocampal and parahippocampal regions, with spread to contralateral regions occurring relatively late (Fig. 6A). We chose to also model a caudoputamen injection site (Fig. 6B) to compare tau pathology spread to previously published data modeling α-synuclein pathology spread. While tau pathology is predicted to spread to some conserved regions, including the substantia nigra and frontal cortical regions, there is also more engagement of thalamic and mesencephalic regions and less engagement of contralateral regions by tau than with α-synuclein pathology.

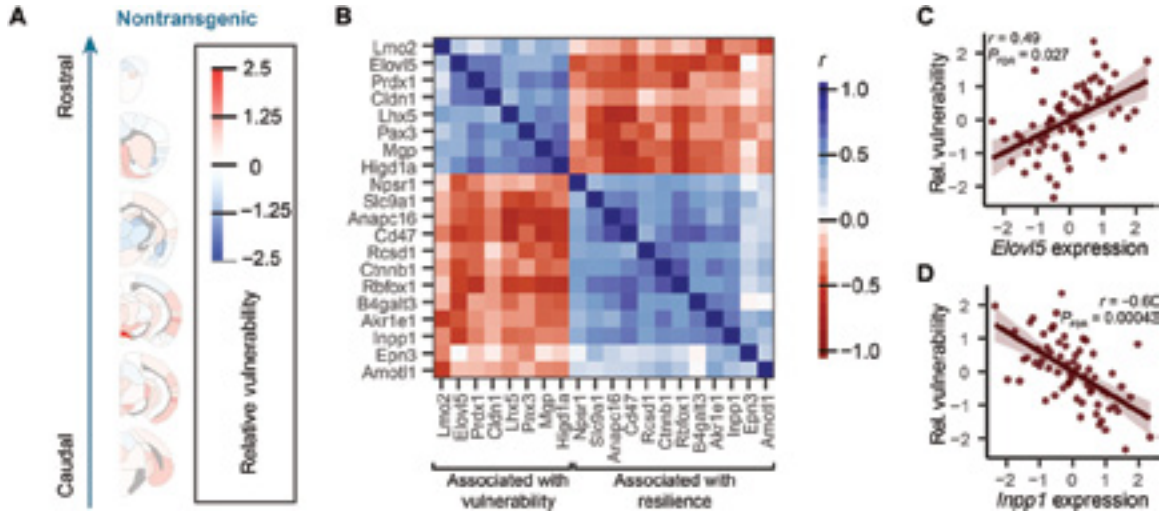


Fig. 5. Model residuals as a predictor of regional vulnerability. (**A**) Residuals between actual tau pathology levels and pathology levels predicted by the bidirectional model were averaged for each region over time and across hemispheres to give an average regional vulnerability to tau pathology. Here, those values are plotted as an anatomical heatmap. (**B**) Gene expression patterns that were statistically significantly associated with regional vulnerability (FDR-corrected $P < 0.05$ cutoff for inclusion) are plotted here with heatmap values indicating the similarity of these gene expression patterns to each other. The genes cluster into two groups, one associated with vulnerability and one associated with resilience, as noted by brackets at the bottom of the plot. (**C**) Normalized relative regional vulnerability is plotted as a function of normalized *Elovl5* expression. (**D**) Normalized relative regional vulnerability is plotted as a function of normalized *Inpp1* expression. For (C) and (D), the solid line represents the line of best fit, and the shaded ribbons represent the 95% prediction intervals. The *r* and *P* values for the Pearson correlation between vulnerability and gene expression are noted on the plots.

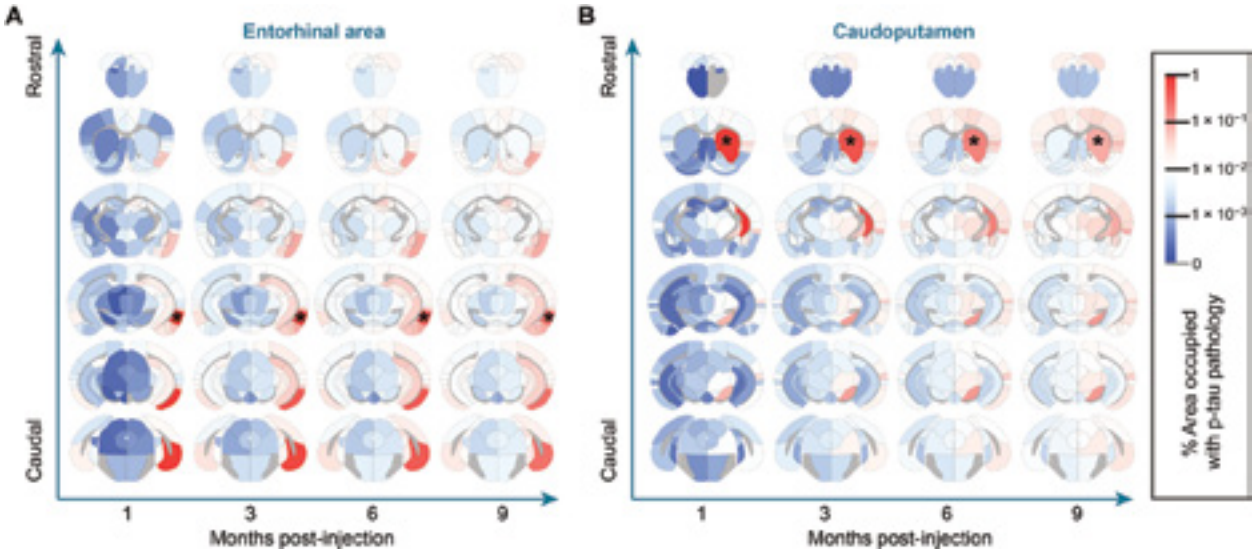


Fig. 6. In silico seeding from alternate sites. Using the diffusion rate constant estimated by model fitting to empirical tau pathology spread, we estimated the distinct spreading patterns that arise after injection into alternate sites in the entorhinal area (**A**) and caudoputamen (**B**). Estimated spread is plotted as a heatmap in anatomical space with blue indicating regions with minimal estimated pathology and red indicating regions with elevated pathology. *Sites of injections.

LRRK2^{G2019S} mice have altered tau pathology patterns

The quantitative pathology-network modeling approach established in NTG mice allows the assessment of dynamics of tau pathology spread. To further investigate the utility of this approach, we performed similar analysis in mice expressing the G2019S mutation in LRRK2. This mutation is the most common cause of familial PD and a common risk factor for idiopathic PD (*30*). Although these patients show similar symptoms to idiopathic PD, neuropathologically, 21 to 54% of patients lack the hallmark α-synuclein Lewy bodies exhibited by patients with idiopathic PD (*31*–*33*). Notably, most of the *LRRK2* mutation carriers exhibit tau pathology (*33*, *34*). To assess alterations in tau pathology distribution and spread related to LRRK2, we performed quantitative analysis of pathology in LRRK2^{G2019S} mice [B6.Cg-Tg(Lrrk2*G2019S)2Yue/J]. We first established that in the absence of pathological tau injection, LRRK2^{G2019S} mice do not accumulate detectable tau pathology up to 12 months of age (fig. S7). Following pathological tau injection, LRRK2^{G2019S} mice accumulate tau pathology in similar regions as NTG mice (Fig. 7A), suggesting that overall spreading is constrained by anatomical connectivity. However, there are clear differences in the regional distribution of tau pathology in LRRK2^{G2019S} mice (Fig. 7B and figs. S8 and S9). In some regions, like the injected iDG and highly connected iSUM, tau pathology is almost identical in NTG and LRRK2^{G2019S} mice (Fig. 7, C and D). In contrast, other regions, especially those that require extended periods of time to exhibit pathology, show elevated pathology in LRRK2^{G2019S} mice (Fig. 7, C and D).

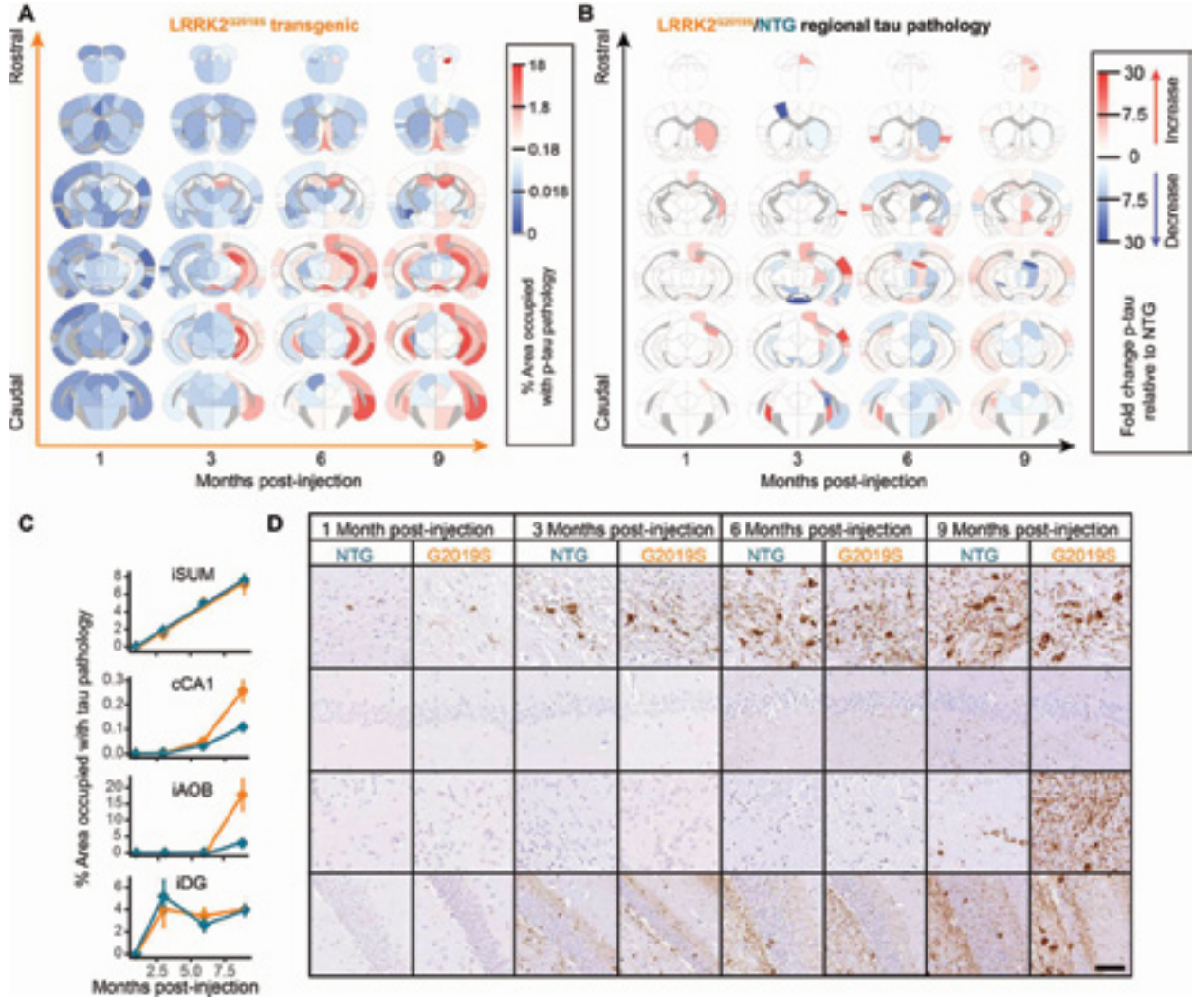


Fig. 7. LRRK2^{G2019S} mice have altered tau pathology patterns. (**A**) Regional pathology measures for LRRK2^{G2019S} mice plotted on anatomical scaffolds as a heatmap, with blue representing minimal pathology, white representing moderate pathology, and red representing substantial pathology. (**B**) The fold change between NTG and LRRK2^{G2019S} mice is plotted on anatomical maps as a heatmap, with blue representing regions with higher pathology in NTG mice and red representing regions with higher pathology in LRRK2^{G2019S} mice. (**C**) The percentage of area occupied with tau pathology is plotted as a function of time for four different regions (iSUM, cCA1, iAOB, and iDG) demonstrating four distinct patterns of pathology propagation in NTG and LRRK2^{G2019S} mice. Some regions, like the iSUM and iDG, show similar tau pathology progression in NTG and LRRK2^{G2019S} mice. In contrast, the cCA1 and iAOB, although similar at 1 and 3 MPI, show enhanced pathology in LRRK2^{G2019S} mice, especially at 9 MPI. (**D**) Representative images of the regions plotted in (C) stained for p-tau and directly adjacent to the plots demonstrating the pathology patterns. Representative images for all regions can be found in the Supplementary Materials. *n* = NTG-1 M: 4, G2019S-1 M: 3, NTG-3 M: 8, G2019S-3 M: 7, NTG-6 M: 6, G2019S-6 M: 5, NTG-9 M: 6, and G2019S-9 M: 4. Scale bar, 50 μm.

These results suggest that LRRK2^{G2019S} alters some aspect of tau pathology spread.

The LRRK2^{G2019S} genetic risk factor alters network dynamics of tau pathology spread

To gain a deeper understanding into what parameters of tau pathology spread may be altered, we assessed spread in LRRK2^{G2019S} mice using network modeling. As suggested by the overall quantitative pathology pattern (Fig. 7A), tau spread can be well explained in LRRK2^{G2019S} mice by anatomical connectivity (Fig. 8A). Similar to NTG mice, tau pathology spread in LRRK2^{G2019S} mice is only moderately well fit at 1 MPI but shows improved fit at 3, 6, and 9 MPI, suggesting that anatomical connectivity is also a major factor driving tau pathology spread in LRRK2^{G2019S} mice. We next sought to explain the regional differences in pathology in LRRK2^{G2019S} mice. We first assessed the relationship of estimated regional vulnerability in NTG mice to the difference in pathology in LRRK2^{G2019S} mice (Fig. 8B). There is a negative correlation between NTG vulnerability and the difference in pathology at 1 and 3 MPI, suggesting that at those time points, there is a shift in pathology in LRRK2^{G2019S} mice from vulnerable regions to more resilient regions. At 6 and 9 MPI, the relationship between NTG vulnerability and pathology difference is diminished. Given the low levels of pathology in NTG and LRRK2^{G2019S} mice at 1 and 3 MPI, the inverse relationship between the difference in pathology in LRRK2^{G2019S} mice and NTG vulnerability may reflect not a shift in vulnerability of regions but early changes in the directionality of pathology movement in the brain.

We next sought to determine whether the difference in regional pathology distribution in LRRK2^{G2019S} mice is related to a difference in spread in anterograde or retrograde directions. To infer the mechanisms of network spread affected by the LRRK2^{G2019S} mutation, we fit the bidirectional model on bootstrapped samples of NTG and LRRK2^{G2019S} mice to obtain distributions of model parameters, namely, the diffusion rate constants and regression weights that measure the relative importance of anterograde and retrograde spread. We observed no statistically significant difference in the overall

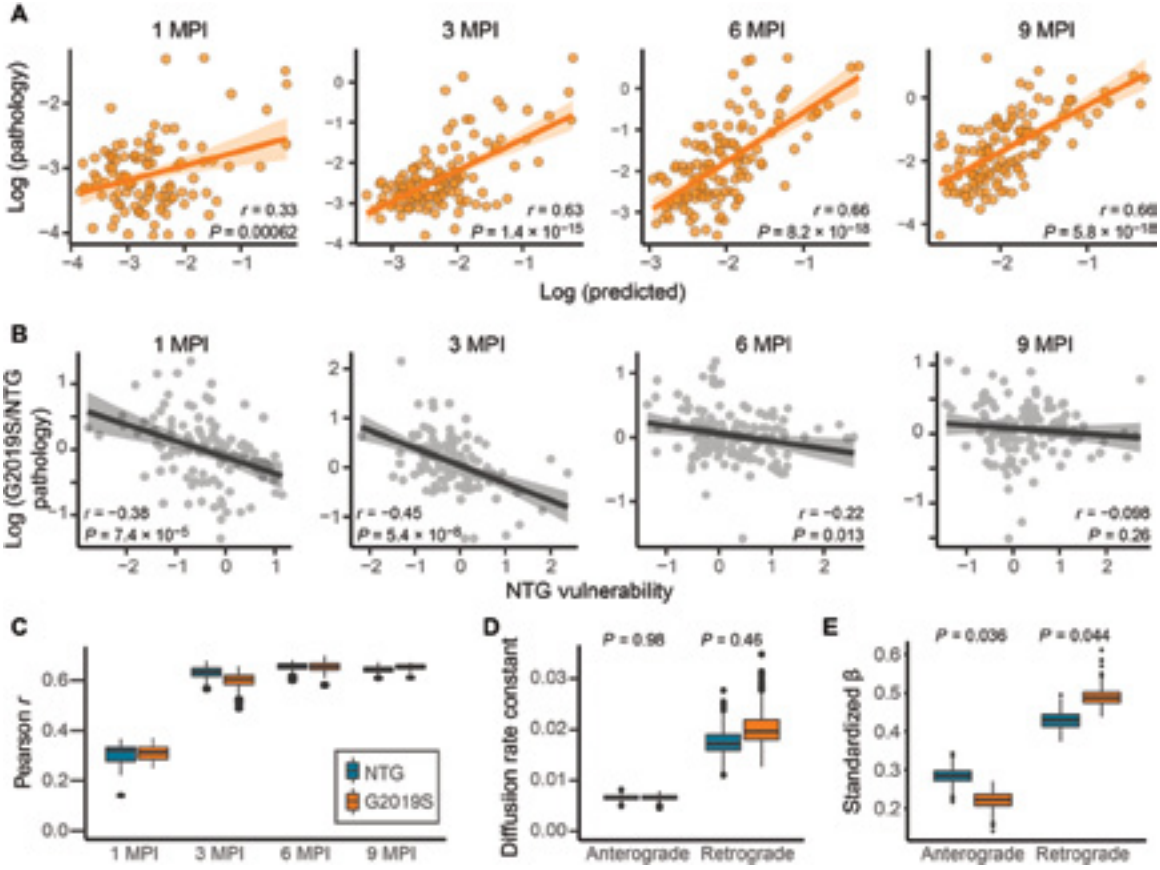


Fig. 8. The LRRK2^{G2019S} genetic risk factor alters network dynamics of tau pathology spread. (**A**) Predictions of regional log tau pathology (*x* axis) in LRRK2^{G2019S} mice from spread models based on retrograde and anterograde anatomical connections, plotted against log actual regional tau pathology values (*y* axis) at 1, 3, 6, and 9 MPI. (**B**) Plots of the NTG vulnerability measure versus log G2019S/NTG pathology, showing a negative correlation between the two measures especially at early time points (1 and 3 MPI) that levels off at later time points (6 and 9 MPI). Solid lines represent the lines of best fit, and the shaded ribbons represent the 95% prediction intervals. The *r* and *P* values for the Pearson correlation between G2019S/NTG pathology ratio and NTG vulnerability are noted on the plots. (**C**) Distributions of model fit (Pearson *r*) for fitting data to bootstrap samples of mice. NTG and G20 do not differ in model fit (nonparametric, two-tailed test). (**D**) Distributions of diffusion rate constants reveal greater intersample variability in retrograde constants compared to anterograde. LRRK2^{G2019S} and NTG do not differ in diffusion rate constants. (**E**) Anterograde and retrograde betas differ between NTG and LRRK2^{G2019S} mice, with LRRK2^{G2019S} preferentially spreading in the retrograde direction. In box plots, box edges represent the 25th and 75th percentiles, the middle line shows the median, and whiskers extend from the box edges to the most extreme data point value that is at most 1.5× IQR. Data beyond the end of the whiskers are plotted individually as dots.

fit of NTG and LRRK2^{G2019S} mice over time (Fig. 8C), suggesting that our network model adequately captures tau spread in LRRK2^{G2019S} mice. In addition, the overall diffusion rate constant along anterograde connections was no different than that along retrograde connections (Fig. 8D), suggesting that the rates of anterograde and retrograde spread are similar. The contribution of anterograde and retrograde connections over time did show differences (Fig. 8E), such that anterograde spread was less important and retrograde spread more important for explaining tau spread patterns in LRRK2^{G2019S} mice. These findings suggest that the LRRK2^{G2019S} mutation may lead to increased shunting of misfolded tau into the already predominant retrograde pathways, which partially explains the differences in pathology patterns observed in these mice.

DISCUSSION

Neurodegenerative diseases progress symptomatically as pathologies appear in previously unaffected regions of the brain. Identification of the neuropathological proteins aggregated in these diseases and subsequent staging studies demonstrated that symptom progression is associated with the presence of aggregated proteins in more and more regions (*7*). The stages of disease suggest that either there is a fine gradient of regional vulnerability such that regions become sequentially affected or pathology can spread through the brain by transcellular means. The latter hypothesis is more parsimonious, and mounting evidence has supported this explanation in recent years. The current study demonstrates using an interdisciplinary approach bridging quantitative pathology and network analysis that tau pathology patterns can be predicted by linear diffusion through the anatomical connectome with modulations of that spread by regional vulnerability.

Our study is congruous with previous work using computational modeling to understand the distribution of tau pathology and related regional atrophy in tauopathies. These studies assessed the utility of computational models to predict semiquantitative regional tau pathology scores in transgenic mice (*35*), global atrophy magnetic resonance imaging in AD and frontotemporal dementia (*12*, *13*), tau positron emission tomography signal in AD (*14*, *15*), or general histopathological patterns of AD (*16*, *17*). While each of these studies used different modeling parameters, each found that a model incorporating spread along anatomical connections was the most efficient and accurate predictor of tau pathology or related atrophy patterns. Our study has now extended these previous efforts in three important ways. First, the current study used tau seeding in an NTG mouse, giving our model a precise spatiotemporal starting point, as defined by the site and time of pathogenic tau injection. Second, we used quantitative measures of mouse tau pathology in 134 regions of the mouse brain, providing several log-fold depth of data throughout the brain. Third, mice were euthanized at four time points, following injection of pathogenic tau, providing pseudo-longitudinal data for model fitting. Our modeling of this data is consistent with previous studies, finding that tau pathology spreads in a pattern that is well described by anatomical connectivity. Further, we were able to delineate the contribution of anterograde and retrograde connections to this spread through a rigorous model comparison approach, assess the kinetics of spread, understand regional vulnerability to pathology, compare regional vulnerability to regional gene expression, and assess the impact of a genetic risk factor on the spread of tau pathology. Future studies could test in silico whether deleting connections would impact spread, which could be validated in vivo with lesioning experiments.

This study has several limitations. One is the resolution of tau pathology data and the reliance on mesoscale connectivity measures to model tau pathology spread. Tau pathology was assessed at a mesoscale to match the mesoscale connectivity atlas generated by the Allen Institute (*27*). However, this scale obscures more granular information, such as the cortical laminar distribution of pathology. The use of the connectivity atlas also imposes technical limitations of connectivity tracing, including a potential bias in cell types infected by tracer viruses, difficulty in distinguishing passing axons from terminals, and possible underestimation of connectivity due to the use of projection volume instead of fluorescence intensity (*27*). Higher-resolution connectivity maps are currently being optimized, and future studies would benefit from higher-resolution pathology maps to match the higher-resolution connectivity maps. A second limitation is the use of a transgenic mouse overexpressing LRRK2. When this study began, this model was desirable because LRRK2 is expressed on the endogenous mouse promoter, ensuring that its regional expression matched that of endogenous LRRK2. However, next-generation models with a knock-in of the LRRK2 mutation are now available (*36*–*38*), and future work should validate the effects of LRRK2 on tau spread in these models.

Despite recent efforts, the neuropathological substrate of *LRRK2* PD has been elusive. While *LRRK2* mutation carriers all have degeneration of substantia nigra neurons, many of them do not accumulate α-synuclein Lewy bodies in the brain (*31*, *32*). This fact has left some question as to the neuropathological substrate of degeneration in these patients. In the search for an alternate pathology, it has now been recognized that most *LRRK2* mutation carriers have tau pathology to varying degrees (*31*, *33*, *34*). Although it is not clear that tau pathology is responsible for dopaminergic neuron death in *LRRK2* mutation carriers, it does raise the possibility that mutations in *LRRK2* can predispose patients to developing tau pathology.

Hyperphosphorylated tau is a well-cataloged feature present in LRRK2 mutant mice (*39*, *40*), flies (*41*), and induced pluripotent stem cell–derived neurons (*42*). It has also been demonstrated that LRRK2 mutations can lead to phosphorylation of tau (*43*–*46*). Although it is not clear that this hyperphosphorylated tau is pathological, it is possible that it represents a pool of tau that is more rapidly recruited upon the introduction of a pathogenic tau seed. The current study was conducted in mice up to 12 months of age, and the LRRK2^{G2019S} mice in this study showed no evidence of phosphorylated tau accumulation without pathogenic tau injection. This observation suggests that the changes in tau pathology spread are likely due not to elevated phosphorylation but rather to alterations in cellular mechanisms such as protein release, internalization, transport, or degradation. This hypothesis is supported by another recent study showing enhanced neuronal transmission of virally expressed tau (*47*). Other studies found that overall tau pathology was not affected in transgenic mice that exhibit broad tau pathology controlled by the transgene promoter (*47*, *48*).

One possible cellular mechanism regulating the change in tau pathology spread in LRRK2^{G2019S} mice is elevated presynaptic vesicle release, a phenomenon that has been observed in LRRK2^{G2019S} mice (*38*). Enhanced synaptic vesicle release is accompanied by enhanced membrane internalization, which could provide an increased opportunity for extracellular tau to be internalized into presynaptic boutons and thereby enhance retrograde transmission of pathology.

Previous research has demonstrated that tau is readily released by neurons (*18*–*20*) in proportion to neuronal activity (*21*, *22*). Future research incorporating functional connectivity strength and tau receptor distribution as regulators of tau spread may help clarify how tau pathology spread is regulated.

In conclusion, the current study demonstrates that tau pathology spreads from an initial injection site through the brain via neuroanatomical connectivity. This spread can be modulated by a genetic risk factor for PD. Future work should substantiate that this alteration is related to LRRK2 kinase activity and explore whether LRRK2 inhibitors would be a viable therapeutic treatment for tauopathies.

MATERIALS AND METHODS

Mice

All housing, breeding, and procedures were performed according to the National Institutes of Health (NIH) Guide for the Care and Use of Experimental Animals and approved by the University of Pennsylvania Institutional Animal Care and Use Committee. C57BL/6 J (NTG, the Jackson laboratory 000664, Research Resource Identifier (RRID): IMSR_JAX:000664) and B6.Cg-Tg(Lrrk2*G2019S)2Yue/J (G2019S, the Jackson laboratory 012467, RRID: IMSR_JAX:012467) mice have been previously described (*49*). The current G2019S bacterial artificial chromosome line was backcrossed to C57BL/6 J mice for >10 generations and bred to homozygosity at loci as determined by quantitative polymerase chain reaction and outbreeding. The expression level of $LRRK2^{G2019S}$ was thereby stabilized in this line of mice. All experiments shown use homozygous $LRRK2^{G2019S}$ mice. Both male (n = 24) and female (n = 19) mice were used and were 3 to 4 months old at the time of injection. No influence of sex was identified in the measures reported in this study.

Human tissue

All procedures were done in accordance with local institutional review board guidelines of the University of Pennsylvania. Written informed consent for autopsy and analysis of tissue sample data was obtained from either patients themselves or their next of kin. All cases used for extraction (table S1) of PHF tau were Braak stage VI and were selected on the basis of a high burden of tau pathology by immunohistochemical staining. Cases used for extraction were balanced by sex (female = 2; male = 2) and were frozen an average of 5 hours postmortem. Differences in sex were not assessed because these cases were only used for protein extraction.

Human brain sequential detergent fractionation

Frozen postmortem human frontal or temporal cortex brain tissue containing abundant tau-positive inclusions was selected for sequential extraction on the basis of immunohistochemistry examination of these samples as described (*50*) using previously established methods. These brains were sequentially extracted with increasing detergent strength as previously described (*9*). After thawing, meninges were removed, and gray matter was carefully separated from white matter. The gray matter was weighed and suspended in nine volumes (w/v) high-salt buffer [10 mM tris-HCl (pH 7.4), 800 mM NaCl, 1 mM EDTA, 2 mM dithiothreitol, 1:1000 protease and 1:100 phosphatase inhibitors, and 1 mM phenylmethylsulfonyl fluoride] with 0.1% sarkosyl and 10% sucrose, followed by homogenization with a dounce homogenizer and centrifugation at 10,000*g* for 10 min at 4°C. The resulting pellet was re-extracted with the same buffer conditions, and the supernatants from all extractions were filtered and pooled.

Additional sarkosyl was added to the pooled supernatant to reach a final concentration of 1%, and the supernatant was nutated for 1 hour at room temperature. The samples were then centrifuged at 300,000*g* for 60 min at 4°C. The pellet, which contains pathological tau, was washed once with phosphate-buffered saline (PBS) and resuspended in 100 μl of PBS per gram of gray matter by passing through a 27-gauge/0.5-inch needle. The pellets were further suspended by brief sonication (QSonica Microson XL-2000; 20 pulses; setting 2; 0.5 s per pulse). The suspension was centrifuged at 100,000*g* for 30 min at 4°C. The pellet was suspended in one-fifth to one-half the pre-centrifugation volume, sonicated briefly (60 to 120 pulses; setting 2; 0.5 s per pulse), and centrifuged at 10,000*g* for 30 min at 4°C. The final supernatant was used for all studies and is referred to as AD PHF tau. All extractions were characterized by Western blotting for tau, sandwich enzyme-linked immunosorbent assay (ELISA) for tau, α-synuclein and Aβ 1–40, Aβ 1–42, and validated by immunocytochemistry in primary neurons from NTG mice. For the extractions used in this study, tau constituted 22.6 to 35.7% of the total protein, while α-synuclein and Aβ constituted 0.011% or less of total protein.

Sandwich ELISA

Characterization of tau, α-synuclein and Aβ 1–40, Aβ 1–42 from AD PHF preparations by sandwich ELISA has been previously described (*9*). Assays were performed on 384-well MaxiSorp clear plates (Thermo Fisher Scientific). Plates were coated with well-characterized capture antibodies (tau: Tau5; α-synuclein: Syn9027; Aβ 1–40/Aβ 1–42: Ban50) at 4°C overnight, washed, and blocked with Block Ace (AbD Serotec) overnight at 4°C. AD PHF preparations were diluted at 1:100 and added to plates alongside serial dilutions of recombinant α-synuclein, recombinant T40, or peptides for Aβ 1–40 and Aβ 1–42 monomeric standards. Plates were incubated overnight at 4°C and then washed and incubated with reporter antibodies (tau: BT2 + BT7; α-synuclein: MJF-R1; Aβ 1–40: BA27; Aβ 1–42: BC05) overnight at 4°C. Plates were washed and incubated with horseradish peroxidase (HRP)–conjugated secondary antibodies for 1 hour at 37°C. Plates were developed with 1-Step Ultra TMB-ELISA Substrate Solution (Thermo Fisher Scientific) for 10 to 15 min. The reaction was quenched with 10% phosphoric acid and read at 450 nm on a plate reader (M5, SpectraMax).

AD PHF stereotaxic injection

All surgery experiments were performed in accordance with protocols approved by the Institutional Animal Care and Use Committee of the University of Pennsylvania. AD PHF tau from individual extractions was vortexed and diluted with Dulbecco's Phosphate Buffered Saline to 0.4 mg/ml. Tau was sonicated (QSonica Microson XL-2000; 60 pulses; setting 1.5; 1 s per pulse). Mice were injected at 3 to 4 months old. Mice were deeply anesthetized with ketamine/xylazine/acepromazine and injected unilaterally by insertion of a single needle into the right forebrain (coordinates: −2.5 mm relative to bregma and +2.0 mm from midline) targeting the hippocampus (2.4 mm beneath the skull) with 1 μg of tau (2.5 μl). The needle was then retracted to 1.4 mm beneath the skull, targeting the overlaying cortex, and another 1 μg of tau (2.5 μl) was injected. The needle was left in place for 2 min following the injection. Injections were performed using a 25-μl syringe (Hamilton, NV) at a rate of 0.4 μl/min.

Immunohistochemistry

After 1 to 9 months, mice were perfused transcardially with PBS; brains were removed and underwent overnight fixation in 70% ethanol in 150 mM NaCl (pH 7.4). After perfusion and fixation, brains were processed into paraffin via sequential dehydration and perfusion with paraffin under vacuum (70% ethanol for 2 hours, 80% ethanol for 1 hour, 95% ethanol for 1 hour, 95% ethanol for 2 hours, three times 100% ethanol for 2 hours, xylene for 30 min, xylene for 1 hour, xylene for 1.5 hours, and three times paraffin for 1 hour at 60°C). Brains were then embedded in paraffin blocks, cut into 6-μm sections, and mounted on glass slides. Slides were then stained using standard immunohistochemistry as described below. Slides were deparaffinized with two sequential 5-min washes in xylenes, followed by 1-min washes in a descending series of ethanols: 100, 100, 95, 80, and 70%. Slides were then incubated in deionized water for 1 min before microwave antigen retrieval in the BioGenex EZ-Retriever System. Slides were incubated in antigen unmasking solution (Vector Laboratories, catalog no. H-3300) and microwaved for 15 min at 95°C. Slides were allowed to cool for 20 min at room temperature and washed in running tap water for 10 min. Slides were incubated in 7.5% hydrogen peroxide in water to quench endogenous peroxidase activity. Slides were washed for 10 min in running tap water, 5 min in 0.1 M tris (diluted from 0.5 M tris made from tris base and concentrated hydrochloric acid to pH 7.6), and then blocked in 0.1 M tris/2% fetal bovine serum (FBS) for 15 min or more. Slides were incubated in primary antibody in 0.1 M tris/2% FBS in a humidified chamber overnight at 4°C. For pathologically phosphorylated tau, pS202/T205 tau (AT8, Thermo Fisher Scientific, catalog no. MN1020) was used at 1:10,000.

Primary antibody was rinsed off with 0.1 M tris, and slides were incubated in 0.1 M tris for 5 min. Primary antibody was detected using a BioGenex Polymer detection kit (catalog no. QD440-XAK) per the manufacturer's protocol as outlined below. Slides were incubated in 50% enhancer solution in 0.1 M tris/2% FBS for 20 min. Enhancer was rinsed off with 0.1 M tris, incubated in 0.1 M tris for 5 min, and incubated in 0.1 M tris/2% FBS for 5 min. Slides were then incubated in 50% poly-HRP in 0.1 M tris/2% FBS for 30 min. Poly-HRP was rinsed off with 0.1 M tris; slides were then incubated for 5 min with 0.1 M tris and then developed with ImmPACT diaminobenzidine (DAB) peroxidase substrate (Vector SK-4105) for 10 min. DAB was rinsed off with 0.1 M tris and incubated in distilled water for 5 min. Slides were then counterstained briefly with Harris hematoxylin (Thermo Fisher Scientific, catalog no. 6765001). Slides were washed in running tap water for 5 min, dehydrated in ascending ethanol for 1 min each (70, 80, 95, 100, and 100%), then washed twice in xylenes for 5 min, and coverslipped in Cytoseal Mounting Media (Thermo Fisher Scientific, catalog no. 23-244-256). Slides were scanned into digital format on a Lamina scanner (PerkinElmer) using a 20× objective (0.8 numerical aperture) into .mrxs files. Digitized slides were then used for quantitative pathology.

Quantitative pathology

All section selection, annotation, and quantification were done blinded to treatment. For quantification of tau pathology, coronal sections were selected to closely match the following coordinates, relative to bregma: 3.20, 0.98, −1.22, −2.92, −3.52, and − 4.48 mm. The digitized images were imported into HALO software to allow annotation and quantification of the percentage area occupied by tau pathology. Standardized annotations were drawn to allow independent quantification of 194 gray matter regions throughout the brain. Brain region annotations were largely based on the ABA (CCF v3, 2017; brain-map.org), although smaller subregions that are difficult to annotate by eye and have minimal pathology were grouped to minimize error and annotation time. For example, thalamic and midbrain nuclei, which accumulate minimal pathology, were grouped into larger regions (fig. S1). Each set of annotations was imported onto the desired section and modified by hand to match the designated brain regions. After annotation, the analysis scripts were applied to the brain to make sure that no nonpathology signal was detected. After annotation of all brains, analysis algorithms were applied to all stained sections, and data analysis measures for each region were recorded.

The total pathology analysis detects total signal above a minimum threshold. Specifically, the analysis included all DAB signal that was above a 0.099 optical density threshold, which was empirically determined to not include any background signal. This signal was then normalized to the total tissue area. A minimal tissue optical density of 0.02 was used to exclude any areas where tissue was split, and a tissue-edge thickness of 25.2 μm was applied to exclude any edge effect staining. Quantitative pathology measures for the 48 regions with substantial pathology were analyzed to determine whether there was any overall genotype effect or a time-dependent effect of genotype on pathology, using a cumulative logistic model. For each region, an interaction between genotype and months was initially investigated via a likelihood ratio test (LRT); this interaction was dropped for models with $P > 0.15$ based on the LRT. LRTs were again performed to determine whether there existed statistically significant genotype difference either across all time points or whether genotype differences changed over time. *P* values from these LRTs were adjusted using the Benjamini-Hochberg FDR to account for multiple testing of 48 regions. Post hoc tests were then performed via the emmeans package in R with, again, an FDR adjustment for multiple testing. Second-generation *P* values were generated, as described in (*51*), with a null interval of odds ratio within 0.9 to 1.1, were also calculated. In general, second-generation *P* values control the familywise error rate and FDR but can often identify additional differences when the number of comparisons is greater than 10. The null interval of 0.9 to 1.1 (i.e., within 10% difference), was chosen as a conservative range.

Computational models of pathological protein spread

Models of linear diffusion along white matter fibers have been used to predict the spread of misfolded α-synuclein in mice (*28*), as well as patterns of atrophy observed in various neurodegenerative diseases (*12*, *52*). In the present work, we extended these models to the spread of tau between anatomically connected brain regions from an injection site in the right hippocampus.

We model pathological spread of tau as a diffusion process on a directed structural brain network $G = \{V, E\}$ whose nodes V are $N_a = 426$ cortical and subcortical gray matter regions and whose edges $e_{ij} \in E$ represent an axonal projection initiating in V_i and terminating in V_j, where $e_{ij} \geq 0$ for all E. Edge strength was quantified by the Allen Brain Institute using measures of fluorescence intensity from anterograde viral tract tracing (*27*). We define the weighted adjacency matrix of G as $\boldsymbol{A} = [\boldsymbol{A}_{ij}]$, such that row i contains the strengths of axonal projections from region i to all other $N_a - 1$ regions. We define the current levels of simulated tau pathology of all N_a nodes at time t as the vector $\boldsymbol{x}(t)$. We make empirical measurements of tau pathology at $t = 1$, 3, 6, and 9 MPI in an $N_c = 134$ region vector $\boldsymbol{y}(t)$, which is a spatially coarse-grained version of $\boldsymbol{x}(t)$. Note that $\boldsymbol{y}(t) = f(\mathbf{x}(\mathrm{t}))$, where f is

a linear transformation that sets each element (region) of $\boldsymbol{y}(t)$ equal to the arithmetic mean of the elements (regions) of $\boldsymbol{x}(t)$ that lie within the regions of $\boldsymbol{y}(t)$. This transformation is needed to avoid quantifying pathology in many of the smaller regions in the N_a-dimensional space used by the Allen Brain Institute (*27*), which are difficult to identify reliably across mice in practice.

In the simplest case of our models of pathological network spread, we simulated the spread of tau pathology throughout the N_a anatomically connected brain regions in $\boldsymbol{A}$ to compute the predicted pathology $\hat{\boldsymbol{y}}(t)$ in N_c empirically assessed brain regions as a function of a set of seed regions $s \in V$ using the form

$$\hat{\boldsymbol{y}}(t) = f(e^{-c_r \boldsymbol{L}_r t}\boldsymbol{x}_o)$$

where the retrograde, out-degree graph Laplacian

$$\boldsymbol{L}_{\boldsymbol{r}_{ij}} = \begin{cases} -\boldsymbol{A}_{ij} \text{ for } i \neq j \\ \sum_{j=1}^{N} \boldsymbol{A}_{ij} \text{ for } i = j \end{cases}$$

$$\boldsymbol{x}_{o_s} = \begin{cases} 0 \text{ for } i \notin s \\ 1 \text{ for } i \in s \end{cases}$$

c_r is a diffusion rate constant representing the global speed of retrograde spread, f is the linear transformation described above that converts the N_a region space to the N_c region space, and t is in units of months. In this manuscript, s contains the ABA regions DG, CA1, CA3, VISam, and RSPagl, to account for experimental variability in targeting a hippocampal injection site. Note that c_r tunes the time scale of the system, which is necessary because of the fact that the units of connection strength are arbitrary relative to the units of pathology. To fit this model, we swept through values of c_r from 10^{-5} to 0.2 and chose the value of c_r that maximized the average Pearson correlation coefficient between $\log_{10} \boldsymbol{y}(t)$ and $\log_{10} \hat{\boldsymbol{y}}(t)$ over $t = [1\ 3\ 6\ 9]$. Empirically, the value of this correlation plateaued for values of c_r larger than 0.2, justifying this upper bound on c_r. Note that $\boldsymbol{L}$ is the out-degree Laplacian, a version of the well-characterized graph Laplacian designed for directed graphs (*53*). Intuitively, this model posits that pathology spreads retrogradely from region i to other regions at a rate proportional to the number of synapses projecting onto i from those regions, while pathology at region i decays as a function of the sum of the strength of projections into region i.

Recent studies by this group (*28*) and others (*54*) have suggested that both anterograde and retrograde spread of pathology contribute to neurodegenerative disease progression. Thus, we expanded the retrograde model described above to a bidirectional model including anterograde spread, using the form

$$\hat{\boldsymbol{y}}(t) = b_o + b_a \log_{10}\left(f(e^{-c_a \boldsymbol{L}_a t}\boldsymbol{x}_o)\right) + b_r \log_{10}(f(e^{-c_r \boldsymbol{L}_r t}\boldsymbol{x}_o)) + \varepsilon(t)$$

where the anterograde, out-degree graph Laplacian

$$\boldsymbol{L}_{\boldsymbol{a}_{ij}} = \begin{cases} -\boldsymbol{A}'_{ij} \text{ for } i \neq j \\ \sum_{j=1}^{N} \boldsymbol{A}'_{ij} \text{ for } i = j \end{cases}$$

c_a is a diffusion rate constant representing the global speed of anterograde spread, b_o is an intercept, b_a is a weight for the importance of anterograde spread, b_r is a weight for the importance of retrograde spread, t is time, and ε is an error term. We used the "optim" function in R to solve for the combination of c_r and c_a that maximizes the average Pearson correlation coefficient between $\log_{10} \boldsymbol{y}(t)$ and $\hat{\boldsymbol{y}}(t)$ over $t = [1\ 3\ 6\ 9]$, with a linear regression inside of the objective function to solve for b_o, b_a, and b_r aggregating across all time points for each c_r and c_a. The linear regression coefficients b_a, and b_r, when standardized, capture the relative importance of anterograde and retrograde spread at time t, respectively, while controlling for the potentially ambiguous overlapping contributions of the two modes of transmission.

We also defined intrinsic regional vulnerability based on $\varepsilon(t)$, the error term in the model above. Intuitively, if $\varepsilon_i(t)$ is large, then this model underpredicted pathology at region i such that region i is more vulnerable to pathology than expected on the basis of bidirectional linear diffusion and vice versa for regions with small values of $\varepsilon(t)$. We hypothesized that both static, intrinsic regional vulnerability and possible time-dependent vulnerability are captured by $\varepsilon(t)$. Therefore, we averaged $\varepsilon(t)$ over hemispheres and over $t = [3\ 6\ 9]$ to capture static, intrinsic regional vulnerability as an $\frac{N_c}{2} \times 1$ vector $\boldsymbol{v_s}$ and average out the effects of both time-dependent vulnerability and unexplainable measurement error. We excluded 1 MPI because pathology at this time point was poorly captured by the spread model, and $\varepsilon(1)$ exhibited a distinct spatial pattern of vulnerability from the patterns observed in $\varepsilon(t)$ at $t = [3\ 6\ 9]$ (fig. S6).

Quantification of model specificity

In a previous study (*28*), we found that the substantial variance explained in misfolded α-synuclein spread by connectome-based linear diffusion models was specific to the use of ipsilateral caudoputamen as the seed site s, which defines the vector $\boldsymbol{x}_o$, over nearly every other region in the connectome. Here, we sought to replicate this finding in the study of tau spread. However, because of the use of multiple seed sites in the present study, additional considerations applied in ensuring a rigorous test of the model's specificity to the experimentally motivated group of seed sites. The five chosen seed sites in s (DG, CA1, CA3, VISam, and RSPagl) were relatively close together (average distance of 1.96 mm), so we wanted to rule out the possibility that (i) choosing multiple random sites would trivially improve model performance and (ii) selecting multiple spatially clustered random seed sites would trivially improve model performance. Thus, we choose 500 random sets of five seed sites $s_{null} \in V$, all of which had an average distance from one another within 1.96 ± 0.196 mm, and fit the bidirectional model detailed above using each set of distance-constrained random points s_{null} to define the initial vector $\boldsymbol{x}_o$. We computed a one-tailed, nonparametric P value for the specificity of s by computing the percentage of times s_{null} yielded a better fit to the observed pathology data than s.

Model comparison

The data presented in the body of the paper use values for the diffusion rate constants, c_r and c_a, and time-dependent weights on retrograde and anterograde spread, b_o, b_a, and b_r, obtained using data from all mice at every time point ($\boldsymbol{y}_{\text{full}}(t)$). To ensure that this approach did not result in overfitting and to rigorously test whether including both anterograde and retrograde connections improved model performance, we randomly sampled without replacement the available mice at each time point to generate $\boldsymbol{y}_{\text{train}}(t)$ and $\boldsymbol{y}_{\text{test}}(t)$ for each time point. These parameters were determined by applying the model fitting procedure described above on $\boldsymbol{y}_{\text{train}}(t)$, and the model was evaluated on the basis of its fit with $\boldsymbol{y}_{\text{test}}(t)$. This process was

repeated 500 times for models based on Euclidean distance, anterograde connections alone, retrograde connections alone, or both anterograde and retrograde connections, generating a distribution of out-of-sample fits for each time point for each model, allowing us to statistically compare the out-of-sample performance between each of the candidate models. Using a similar approach, we generated distributions of model parameters for the bidirectional model using bootstrapped samples of NTG and LRRK2 G2019S mice.

Network null models

To ensure that our results were specific to the retrograde spread of misfolded synuclein along neuronal processes, we repeated our analyses using several network null models. To demonstrate a general specificity of the model for the topology of the anatomic connectome represented by $\boldsymbol{A}$, we carried out a procedure that rewires the edges of G while exactly preserving either the out-degree or the in-degree sequence, i.e.,

$$\sum_{j=1}^{N} \boldsymbol{B}_{ij} \text{ and } \sum_{i=1}^{N} \boldsymbol{B}_{ij}, \text{ where } \boldsymbol{B}_{ij} = \begin{cases} 1 \text{ for } \boldsymbol{A}_{ij} > 0 \\ 0 \text{ otherwise} \end{cases}$$

This rewiring approach tests whether the model fit is due to a relatively basic structural property of the graph, i.e., degree, as opposed to unique, higher-order topological features of the synaptic connectome. Last, we tested the null model that spread of misfolded protein occurs simply because of diffusion through tissue based on closeness in Euclidean space. To test this model, we reconstructed $\boldsymbol{A} = [\boldsymbol{A}_{ij}]$ such that the edges represented the inverse Euclidean distance between region i and region j. We used the initial procedure for retrograde model fitting to test this model, because Euclidean distance is symmetric and thus the bidirectional model cannot be applied.

Assessment of regional gene expression patterns

To assess the cellular and molecular characteristics of intrinsically vulnerable or resilient regions, we compared the spatial alignment between our $\frac{N_c}{2} \times 1$ vector of regional vulnerability measurements, $\boldsymbol{v_s}$, with microarray gene expression levels obtained from the Allen Mouse Brain Atlas (brain-map.org). After applying a previously validated quality control approach to hone in on genes with the most reliable expression measurements (*55*), we obtained $\boldsymbol{G}$, an $\frac{N_c}{2} \times 4277$ matrix of gene expression values across each brain region. Because of the non-normality of both $\boldsymbol{v_s}$ and gene expression in $\boldsymbol{G}$, we applied a rank inverse normalization transform (*56*) to each column of $\boldsymbol{G}$ to control type I error and maximize statistical power. Next, we computed 4277 spatial Pearson correlations between $\boldsymbol{v_s}$ and each column of $\boldsymbol{G}$. We performed multiple comparisons correction by controlling the FDR at $q < 0.05$.

Quantification and statistical analysis

The number of animals analyzed in each experiment, the statistical analysis performed, and the P values for all results <0.05 are reported in the figure legends. In vivo pathological spread data were analyzed, and all computations were performed in R (www.R-project.org/) (*57*) as described.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/24/eabg6677/DC1

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. GBD 2016 Parkinson's Disease Collaborators, Global, regional, and national burden of Parkinson's disease, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **17**, 939–953 (2018).
2. GBD 2016 Dementia Collaborators, Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **18**, 88–106 (2019).
3. J. L. Robinson, E. B. Lee, S. X. Xie, L. Rennert, E. R. Suh, C. Bredenberg, C. Caswell, V. M. van Deerlin, N. Yan, A. Yousef, H. I. Hurtig, A. Siderowf, M. Grossman, C. T. McMillan, B. Miller, J. E. Duda, D. J. Irwin, D. Wolk, L. Elman, L. McCluskey, A. Chen-Plotkin, D. Weintraub, S. E. Arnold, J. Brettschneider, V. M. Y. Lee, J. Q. Trojanowski, Neurodegenerative disease concomitant proteinopathies are prevalent, age-related and APOE4-associated. *Brain* **141**, 2181–2193 (2018).
4. D. G. Coughlin, H. I. Hurtig, D. J. Irwin, Pathological influences on clinical heterogeneity in Lewy body diseases. *Mov. Disord.* **35**, 5–19 (2020).
5. D. Coughlin, S. X. Xie, M. Liang, A. Williams, C. Peterson, D. Weintraub, C. T. McMillan, D. A. Wolk, R. S. Akhtar, H. I. Hurtig, H. B. Coslett, R. H. Hamilton, A. D. Siderowf, J. E. Duda, K. Rascovsky, E. B. Lee, V. M.-Y. Lee, M. Grossman, J. Q. Trojanowski, D. J. Irwin, Cognitive and pathological influences of tau pathology in Lewy body disorders. *Ann. Neurol.* **85**, 259–271 (2019).
6. D. J. Irwin, M. Grossman, D. Weintraub, H. I. Hurtig, J. E. Duda, S. X. Xie, E. B. Lee, V. M. van Deerlin, O. L. Lopez, J. K. Kofler, P. T. Nelson, G. A. Jicha, R. Woltjer, J. F. Quinn, J. Kaye, J. B. Leverenz, D. Tsuang, K. Longfellow, D. Yearout, W. Kukull, C. D. Keene, T. J. Montine, C. P. Zabetian, J. Q. Trojanowski, Neuropathological and genetic correlates of survival and dementia onset in synucleinopathies: A retrospective analysis. *Lancet Neurol.* **16**, 55–65 (2017).
7. H. Braak, E. Braak, Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* **82**, 239–259 (1991).
8. H. Braak, I. Alafuzoff, T. Arzberger, H. Kretzschmar, K. Del Tredici, Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol.* **112**, 389–404 (2006).
9. J. L. Guo, S. Narasimhan, L. Changolkar, Z. He, A. Stieber, B. Zhang, R. J. Gathagan, M. Iba, J. D. McBride, J. Q. Trojanowski, V. M. Y. Lee, Unique pathological tau conformers from Alzheimer's brains transmit tau pathology in nontransgenic mice. *J. Exp. Med.* **213**, 2635–2654 (2016).
10. S. Narasimhan, J. L. Guo, L. Changolkar, A. Stieber, J. D. McBride, L. V. Silva, Z. He, B. Zhang, R. J. Gathagan, J. Q. Trojanowski, V. M. Y. Lee, Pathological tau strains from human brains recapitulate the diversity of tauopathies in nontransgenic mouse brain. *J. Neurosci.* **37**, 11406–11423 (2017).
11. B. B. McAllister, S. G. Lacoursiere, R. J. Sutherland, M. H. Mohajerani, Intracerebral seeding of amyloid-β and tau pathology in mice: Factors underlying prion-like spreading and comparisons with α-synuclein. *Neurosci. Biobehav. Rev.* **112**, 1–27 (2020).
12. A. Raj, A. Kuceyeski, M. Weiner, A network diffusion model of disease progression in dementia. *Neuron* **73**, 1204–1215 (2012).
13. J. A. Brown, J. Deng, J. Neuhaus, I. J. Sible, A. C. Sias, S. E. Lee, J. Kornak, G. A. Marx, A. M. Karydas, S. Spina, L. T. Grinberg, G. Coppola, D. H. Geschwind, J. H. Kramer, M. L. Gorno-Tempini, B. L. Miller, H. J. Rosen, W. W. Seeley, Patient-tailored, connectivity-based forecasts of spreading brain atrophy. *Neuron* **104**, 856–868.e5 (2019).
14. R. Ossenkoppele, L. Iaccarino, D. R. Schonhaut, J. A. Brown, R. la Joie, J. P. O'Neil, M. Janabi, S. L. Baker, J. H. Kramer, M. L. Gorno-Tempini, B. L. Miller, H. J. Rosen, W. W. Seeley, W. J. Jagust, G. D. Rabinovici, Tau covariance patterns in Alzheimer's disease patients match intrinsic connectivity networks in the healthy brain. *NeuroImage Clin.* **23**, 101848 (2019).
15. J. W. Vogel, Y. Iturria-Medina, O. T. Strandberg, R. Smith, E. Levitis, A. C. Evans, O. Hansson; Alzheimer's Disease Neuroimaging Initiative; Swedish BioFinder Study, Spread of pathological tau proteins through communicating neurons in human Alzheimer's disease. *Nat. Commun.* **11**, 2612 (2020).
16. S. Fornari, A. Schäfer, M. Jucker, A. Goriely, E. Kuhl, Prion-like spreading of Alzheimer's disease within the brain's connectome. *J. R. Soc. Interface* **16**, 20190356 (2019).
17. J. J. Weickenmeier, M. Jucker, A. Goriely, E. Kuhl, A physics-based model explains the prion-like features of neurodegeneration in Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis. *J. Mech. Phys. Solids* **124**, 264–281 (2019).
18. S. Saman, W. H. Kim, M. Raya, Y. Visnick, S. Miro, S. Saman, B. Jackson, A. C. McKee, V. E. Alvarez, N. C. Y. Lee, G. F. Hall, Exosome-associated tau is secreted in tauopathy models and is selectively phosphorylated in cerebrospinal fluid in early Alzheimer disease. *J. Biol. Chem.* **287**, 3842–3849 (2012).
19. Y. Wang, V. Balaji, S. Kaniyappan, L. Krüger, S. Irsen, K. Tepper, R. R. Chandupatla, W. Maetzler, A. Schneider, E. Mandelkow, E. M. Mandelkow, The release and trans-synaptic transmission of Tau via exosomes. *Mol. Neurodegener.* **12**, 5 (2017).
20. T. Katsinelos, M. Zeitler, E. Dimou, A. Karakatsani, H. M. Müller, E. Nachman, J. P. Steringer, C. Ruiz de Almodovar, W. Nickel, T. R. Jahn, Unconventional secretion mediates the trans-cellular spreading of tau. *Cell Rep.* **23**, 2039–2055 (2018).

21. J. W. Wu, S. A. Hussaini, I. M. Bastille, G. A. Rodriguez, A. Mrejeru, K. Rilett, D. W. Sanders, C. Cook, H. Fu, R. A. C. M. Boonen, M. Herman, E. Nahmani, S. Emrani, Y. H. Figueroa, M. I. Diamond, C. L. Clelland, S. Wray, K. E. Duff, Neuronal activity enhances tau propagation and tau pathology in vivo. *Nat. Neurosci.* **19**, 1085–1092 (2016).
22. J. K. Holth, S. K. Fritschi, C. Wang, N. P. Pedersen, J. R. Cirrito, T. E. Mahan, M. B. Finn, M. Manis, J. C. Geerling, P. M. Fuller, B. P. Lucey, D. M. Holtzman, The sleep-wake cycle regulates brain interstitial fluid tau in mice and CSF tau in humans. *Science* **363**, 880–884 (2019).
23. B. B. Holmes, S. L. DeVos, N. Kfoury, M. Li, R. Jacks, K. Yanamandra, M. O. Ouidja, F. M. Brodsky, J. Marasa, D. P. Bagchi, P. T. Kotzbauer, T. M. Miller, D. Papy-Garcia, M. I. Diamond, Heparan sulfate proteoglycans mediate internalization and propagation of specific proteopathic seeds. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E3138–E3147 (2013).
24. M. Usenovic, S. Niroomand, R. E. Drolet, L. Yao, R. C. Gaspar, N. G. Hatcher, J. Schachter, J. J. Renger, S. Parmentier-Batteur, Internalized tau oligomers cause neurodegeneration by inducing accumulation of pathogenic tau in human neurons derived from induced pluripotent stem cells. *J. Neurosci.* **35**, 14234–14250 (2015).
25. J. N. Rauch, J. J. Chen, A. W. Sorum, G. M. Miller, T. Sharf, S. K. See, L. C. Hsieh-Wilson, M. Kampmann, K. S. Kosik, Tau internalization is regulated by 6-*O* sulfation on heparan sulfate proteoglycans (HSPGs). *Sci. Rep.* **8**, 6382 (2018).
26. J. N. Rauch, G. Luna, E. Guzman, M. Audouard, C. Challis, Y. E. Sibih, C. Leshuk, I. Hernandez, S. Wegmann, B. T. Hyman, V. Gradinaru, M. Kampmann, K. S. Kosik, LRP1 is a master regulator of tau uptake and spread. *Nature* **580**, 381–385 (2020).
27. S. W. Oh, J. A. Harris, L. Ng, B. Winslow, N. Cain, S. Mihalas, Q. Wang, C. Lau, L. Kuan, A. M. Henry, M. T. Mortrud, B. Ouellette, T. N. Nguyen, S. A. Sorensen, C. R. Slaughterbeck, W. Wakeman, Y. Li, D. Feng, A. Ho, E. Nicholas, K. E. Hirokawa, P. Bohn, K. M. Joines, H. Peng, M. J. Hawrylycz, J. W. Phillips, J. G. Hohmann, P. Wohnoutka, C. R. Gerfen, C. Koch, A. Bernard, C. Dang, A. R. Jones, H. Zeng, A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).
28. M. X. Henderson, E. J. Cornblath, A. Darwich, B. Zhang, H. Brown, R. J. Gathagan, R. M. Sandler, D. S. Bassett, J. Q. Trojanowski, V. M. Y. Lee, Spread of α-synuclein pathology through the brain connectome is modulated by selective vulnerability and predicted by network analysis. *Nat. Neurosci.* **22**, 1248–1257 (2019).
29. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B. Methodol.* **57**, 289–300 (1995).
30. D. G. Healy, M. Falchi, S. S. O'Sullivan, V. Bonifati, A. Durr, S. Bressman, A. Brice, J. Aasly, C. P. Zabetian, S. Goldwurm, J. J. Ferreira, E. Tolosa, D. M. Kay, C. Klein, D. R. Williams, C. Marras, A. E. Lang, Z. K. Wszolek, J. Berciano, A. H. Schapira, T. Lynch, K. P. Bhatia, T. Gasser, A. J. Lees, N. W. Wood; International LRRK2 Consortium, Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: A case-control study. *Lancet Neurol.* **7**, 583–590 (2008).
31. M. Poulopoulos, O. A. Levy, R. N. Alcalay, The neuropathology of genetic Parkinson's disease. *Mov. Disord.* **27**, 831–842 (2012).
32. L. V. Kalia, A. E. Lang, L. N. Hazrati, S. Fujioka, Z. K. Wszolek, D. W. Dickson, O. A. Ross, V. M. van Deerlin, J. Q. Trojanowski, H. I. Hurtig, R. N. Alcalay, K. S. Marder, L. N. Clark, C. Gaig, E. Tolosa, J. Ruiz-Martínez, J. F. Marti-Masso, I. Ferrer, A. López de Munain, S. M. Goldman, B. Schüle, J. W. Langston, J. O. Aasly, M. T. Giordana, V. Bonifati, A. Puschmann, M. Canesi, G. Pezzoli, A. Maues de Paula, K. Hasegawa, C. Duyckaerts, A. Brice, A. J. Stoessl, C. Marras, Clinical correlations with Lewy body pathology in LRRK2-related Parkinson disease. *JAMA Neurol.* **72**, 100–105 (2015).
33. M. X. Henderson, M. Sengupta, J. Q. Trojanowski, V. M. Y. Lee, Alzheimer's disease tau is a prominent pathology in LRRK2 Parkinson's disease. *Acta Neuropathol. Commun.* **7**, 183 (2019).
34. C. Blauwendraat, O. Pletnikova, J. T. Geiger, N. A. Murphy, Y. Abramzon, G. Rudow, A. Mamais, M. S. Sabir, B. Crain, S. Ahmed, L. S. Rosenthal, C. C. Bakker, F. Faghri, R. Chia, J. Ding, T. M. Dawson, A. Pantelyat, M. S. Albert, M. A. Nalls, S. M. Resnick, L. Ferrucci, M. R. Cookson, A. E. Hillis, J. C. Troncoso, S. W. Scholz, Genetic analysis of neurodegenerative diseases in a pathology cohort. *Neurobiol. Aging* **76**, 214.e1–214.e9 (2019).
35. C. Mezias, E. LoCastro, C. Xia, A. Raj, Connectivity, not region-intrinsic properties, predicts regional vulnerability to progressive tau pathology in mouse models of disease. *Acta Neuropathol. Commun.* **5**, 61 (2017).
36. J. Schapansky, S. Khasnavis, M. P. DeAndrade, J. D. Nardozzi, S. R. Falkson, J. D. Boyd, J. B. Sanderson, T. Bartels, H. L. Melrose, M. J. LaVoie, Familial knockin mutation of LRRK2 causes lysosomal dysfunction and accumulation of endogenous insoluble α-synuclein in neurons. *Neurobiol. Dis.* **111**, 26–35 (2018).
37. B. A. Matikainen-Ankney, N. Kezunovic, R. E. Mesias, Y. Tian, F. M. Williams, G. W. Huntley, D. L. Benson, Altered development of synapse structure and function in striatum caused by Parkinson's disease-linked LRRK2-G2019S mutation. *J. Neurosci.* **36**, 7128–7141 (2016).
38. D. A. Beccano-Kelly, N. Kuhlmann, I. Tatarnikov, M. Volta, L. N. Munsie, P. Chou, L.-P. Cao, H. Han, L. Tapia, M. J. Farrer, A. J. Milnerwood, Synaptic function is modulated by LRRK2 and glutamate release is increased in cortical neurons of G2019S LRRK2 knock-in mice. *Front. Cell. Neurosci.* **8**, 301 (2014).
39. Y. Li, W. Liu, T. F. Oo, L. Wang, Y. Tang, V. Jackson-Lewis, C. Zhou, K. Geghman, M. Bogdanov, S. Przedborski, M. F. Beal, R. E. Burke, C. Li, Mutant LRRK2(R1441G) BAC transgenic mice recapitulate cardinal features of Parkinson's disease. *Nat. Neurosci.* **12**, 826–828 (2009).
40. H. L. Melrose, J. C. Dächsel, B. Behrouz, S. J. Lincoln, M. Yue, K. M. Hinkle, C. B. Kent, E. Korvatska, J. P. Taylor, L. Witten, Y. Q. Liang, J. E. Beevers, M. Boules, B. N. Dugger, V. A. Serna, A. Gaukhman, X. Yu, M. Castanedes-Casey, A. T. Braithwaite, S. Ogholikhan, N. Yu, D. Bass, G. Tyndall, G. D. Schellenberg, D. W. Dickson, C. Janus, M. J. Farrer, Impaired dopaminergic neurotransmission and microtubule-associated protein tau alterations in human LRRK2 transgenic mice. *Neurobiol. Dis.* **40**, 503–517 (2010).
41. C. H. Lin, P. I. Tsai, R. M. Wu, C. T. Chien, LRRK2 G2019S mutation induces dendrite degeneration through mislocalization and phosphorylation of tau by recruiting autoactivated GSK3ss. *J. Neurosci.* **30**, 13138–13149 (2010).
42. E. Ohta, T. Nihira, A. Uchino, Y. Imaizumi, Y. Okada, W. Akamatsu, K. Takahashi, H. Hayakawa, M. Nagai, M. Ohyama, M. Ryo, M. Ogino, S. Murayama, A. Takashima, K. Nishiyama, Y. Mizuno, H. Mochizuki, F. Obata, H. Okano, I2020T mutant LRRK2 iPSC-derived neurons in the Sagamihara family exhibit increased tau phosphorylation through the AKT/GSK-3β signaling pathway. *Hum. Mol. Genet.* **24**, 4879–4900 (2015).
43. F. Kawakami, T. Yabata, E. Ohta, T. Maekawa, N. Shimada, M. Suzuki, H. Maruyama, T. Ichikawa, F. Obata, LRRK2 phosphorylates tubulin-associated tau but not the free molecule: LRRK2-mediated regulation of the tau-tubulin association and neurite outgrowth. *PLOS ONE* **7**, e30834 (2012).
44. R. M. Bailey, J. P. Covy, H. L. Melrose, L. Rousseau, R. Watkinson, J. Knight, S. Miles, M. J. Farrer, D. W. Dickson, B. I. Giasson, J. Lewis, LRRK2 phosphorylates novel tau epitopes and promotes tauopathy. *Acta Neuropathol.* **126**, 809–827 (2013).
45. M. R. Shanley, D. Hawley, S. Leung, N. F. Zaidi, R. Dave, K. A. Schlosser, R. Bandopadhyay, S. A. Gerber, M. Liu, LRRK2 facilitates tau phosphorylation through strong interaction with tau and cdk5. *Biochemistry* **54**, 5198–5208 (2015).
46. M. Hamm, R. Bailey, G. Shaw, S. H. Yen, J. Lewis, B. I. Giasson, Physiologically relevant factors influence tau phosphorylation by leucine-rich repeat kinase 2. *J. Neurosci. Res.* **93**, 1567–1580 (2015).
47. A. P. T. Nguyen, G. Daniel, P. Valdés, M. S. Islam, B. L. Schneider, D. J. Moore, G2019S LRRK2 enhances the neuronal transmission of tau in the mouse brain. *Hum. Mol. Genet.* **27**, 120–134 (2018).
48. F. Mikhail, N. Calingasan, L. Parolari, A. Subramanian, L. Yang, M. Flint Beal, Lack of exacerbation of neurodegeneration in a double transgenic mouse model of mutant LRRK2 and tau. *Hum. Mol. Genet.* **24**, 3545–3556 (2015).
49. X. Li, J. C. Patel, J. Wang, M. V. Avshalumov, C. Nicholson, J. D. Buxbaum, G. A. Elder, M. E. Rice, Z. Yue, Enhanced striatal dopamine transmission and motor performance with LRRK2 overexpression in mice is eliminated by familial Parkinson's disease mutation G2019S. *J. Neurosci.* **30**, 1788–1797 (2010).
50. D. J. Irwin, M. T. White, J. B. Toledo, S. X. Xie, J. L. Robinson, V. van Deerlin, V. M. Y. Lee, J. B. Leverenz, T. J. Montine, J. E. Duda, H. I. Hurtig, J. Q. Trojanowski, Neuropathologic substrates of Parkinson disease dementia. *Ann. Neurol.* **72**, 587–598 (2012).
51. J. D. Blume, R. A. Greevy, V. F. Welty, J. R. Smith, W. D. Dupont, An introduction to second-generation *P* values. *Am. Stat.* **73**, 157–167 (2019).
52. S. Pandya, C. Mezias, A. Raj, Predictive model of spread of progressive supranuclear palsy using directional network diffusion. *Front. Neurol.* **8**, 692 (2017).
53. W. H. R. Chan, M. Wildemeersch, T. Q. S. Quek, Characterization and control of diffusion processes in multi-agent networks. arXiv:1508.06738 [cs.SI] (27 August 2015).
54. C. Mezias, N. Rey, P. Brundin, A. Raj, Neural connectivity predicts spreading of alpha-synuclein pathology in fibril-injected mouse models: Involvement of retrograde and anterograde axonal propagation. *Neurobiol. Dis.* **134**, 104623 (2019).
55. B. D. Fulcher, J. D. Murray, V. Zerbi, X. J. Wang, Multimodal gradients across mouse cortex. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4689–4695 (2019).
56. P. L. Auer, A. P. Reiner, S. M. Leal, The effect of phenotypic outliers and non-normality on rare-variant association testing. *Eur. J. Hum. Genet.* **24**, 1188–1194 (2016).
57. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020); www.R-project.org/.

Acknowledgments: We would like to thank members of the laboratory for feedback in developing this manuscript and the Van Andel Research Institute Bioinformatics and Biostatistics Core, especially Z. Madaj, for assistance with statistical analysis of tau pathology. **Funding:** This study was supported by the Michael J. Fox Foundation for Parkinson's Research grant 16879 (to M.X.H.); NIH grants T32-AG000255 (to V.M.Y.L.), P30-AG10124 (to J.Q.T.), U19-AG062418 (to J.Q.T.), P50-NS053488 (to J.Q.T.), R01-NS099348 (to D.S.B.), F30 MH118871-01 (to E.J.C.); and NSF grants PHY-1554488 (to D.S.B.) and BCS-1631550 (to D.S.B.). D.S.B. also acknowledges support from the John D. and Catherine T. MacArthur Foundation, the ISI Foundation, the Alfred P. Sloan Foundation, and the Paul G. Allen Foundation. **Author contributions:**

Conceptualization: M.X.H. and V.M.Y.L. Methodology: M.X.H., E.J.C., and V.M.Y.L. Software: E.J.C. and D.S.B. Formal analysis: M.X.H. and E.J.C. Investigation: M.X.H., E.J.C., H.L.L., L.C., B.Z., H.J.B., R.J.G., and M.F.O. Resources: L.C. Writing—original draft, M.X.H. and E.J.C. Writing—review and editing, all. Visualization: M.X.H. and E.J.C. Supervision: M.X.H., D.S.B., J.Q.T., and V.M.Y.L. Funding acquisition: M.X.H., D.S.B., J.Q.T., and V.M.Y.L. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Primary data and code used to generate the spread modeling are available on GitHub (https://github.com/ejcorn/tau-spread).

Submitted 20 January 2021
Accepted 21 April 2021
Published 9 June 2021
10.1126/sciadv.abg6677

Citation: E. J. Cornblath, H. L. Li, L. Changolkar, B. Zhang, H. J. Brown, R. J. Gathagan, M. F. Olufemi, J. Q. Trojanowski, D. S. Bassett, V. M. Y. Lee, M. X. Henderson, Computational modeling of tau pathology spread reveals patterns of regional vulnerability and the impact of a genetic risk factor. *Sci. Adv.* **7**, eabg6677 (2021).

CANCER

Concordance of hydrogen peroxide–induced 8-oxo-guanine patterns with two cancer mutation signatures of upper GI tract tumors

Seung-Gi Jin, Yingying Meng, Jennifer Johnson, Piroska E. Szabó, Gerd P. Pfeifer*

Oxidative DNA damage has been linked to inflammation, cancer, and aging. Here, we have mapped two types of oxidative DNA damage, oxidized guanines produced by hydrogen peroxide and oxidized thymines created by potassium permanganate, at a single-base resolution. 8-Oxo-guanine occurs strictly dependent on the G/C sequence context and shows a pronounced peak at transcription start sites (TSSs). We determined the trinucleotide sequence pattern of guanine oxidation. This pattern shows high similarity to the cancer-associated single-base substitution signatures SBS18 and SBS36. SBS36 is found in colorectal cancers that carry mutations in *MUTYH*, encoding a repair enzyme that operates on 8-oxo-guanine mispairs. SBS18 is common in inflammation-associated upper gastrointestinal tract tumors including esophageal and gastric adenocarcinomas. Oxidized thymines induced by permanganate occur with a distinct dinucleotide specificity, 5′T-A/C, and are depleted at the TSS. Our data suggest that two cancer mutational signatures, SBS18 and SBS36, are caused by reactive oxygen species.

INTRODUCTION

Reactive oxygen species (ROS) from internal and external sources represent a threat to genome integrity. Different types of ROS are generated during cellular metabolism or from exposure to chemicals, air pollution, ultraviolet (UV) radiation, or diet (*1*–*6*). These ROS include hydrogen peroxide (H_2O_2) as a common oxidizing molecule present in cells and molecules derived from H_2O_2 in the presence of transition metals, such as hydroxyl radicals (•OH). Superoxide anion ($O_2^{\bullet-}$) produced by electron leakage from the electron transport chain or by myeloperoxidase during inflammatory processes is converted to H_2O_2 by superoxide dismutase (*7*). Another relevant type of ROS is singlet oxygen ($^1O_2$) formed by UVA radiation through excitation of endogenous photosensitizers (*2*). Reactive nitrogen species are produced during inflammatory processes (*8*). These molecules react with guanine to form 8-oxo-7,8-dihydro-2′-deoxyguanosine (8-oxo-dG). Guanine has the lowest oxidation potential compared to the other DNA bases and is therefore oxidized preferentially. 8-Oxoguanine (8-oxo-G) is the most prevalent DNA base oxidation product, although more than 20 oxidatively damaged DNA bases have been identified (*1*–*3*, *5*). In addition to direct oxidation of guanine, 8-oxo-dG in DNA may arise from oxidation of the nucleotide pool, i.e., by formation of 8-oxo-dGTP, which is then incorporated into DNA during replication (*9*). Damage of pyrimidines is a less efficient reaction produced by ROS. One exception is the oxidizing agent potassium permanganate ($KMnO_4$), which readily oxidizes thymine to form mostly thymine glycol (*10*, *11*).

8-oxo-G is repaired primarily by base excision repair initiated by the 8-oxoguanine DNA glycosylase OGG1 (*1*). When this lesion is not repaired and persists during replication, DNA polymerases may incorporate cytosine or adenine across from 8-oxo-dG. A specialized repair DNA glycosylase, MUTYH (MYH), is present in mammalian cells to remove the misincorporated adenine from 8-oxo-G/A mispairs (*12*). When these repair pathways fail, the most common mutation induced by 8-oxo-G is the G to T transversion (*13*, *14*). Guanine oxidation may contribute to carcinogenesis. *MUTYH* mutations have been associated with inherited colorectal polyposis and colorectal cancer (*15*, *16*) and possibly other cancers (*17*). In mouse models, combined deficiency of *Myh* (encoding MUTYH) and *Ogg1* (encoding the OGG1 repair enzyme) predisposes the animals to lung and ovarian tumors and lymphomas. In these mice, 75% of the lung tumors carried G to T mutations in the *Kras* gene (*18*).

It has been difficult to unambiguously link ROS and oxidative DNA damage to human cancer. The production of ROS is a well-known process occurring in the setting of inflammation. An estimated 25% of all human cancers have been linked to chronic inflammation, which acts as a tumor-predisposing condition, and many cancer risk factors have inflammation as a common mode of action (*19*). These malignancies include cancers associated with infections (for example, liver cancers and hepatitis viruses, cervical

Department of Epigenetics, Van Andel Institute, Grand Rapids, MI 49503, USA.
*Corresponding author. Email: gerd.pfeifer@vai.org

cancers and head and neck tumors and human papilloma virus, and gastric cancers and *Helicobacter pylori*) or inflammatory conditions [liver cancers and alcoholic and nonalcoholic steatohepatitis (fatty liver disease), esophageal cancers and Barret's esophagus, intestinal tumors linked to inflammatory bowel disease, and several others].

As part of the Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium, more than 84 million mutations from whole cancer genomes and exomes were used to establish about 50 single-base substitution (SBS) signatures (*20*). Some signatures occur in most cancer types and in most individual tumors analyzed. An example is SBS1, with typical C to T mutations at CpG dinucleotides. SBS1 is thought to be derived from spontaneous hydrolytic deamination of 5-methylcytosine at methylated CpG sites in the genome (*21*). On the other hand, many signatures are quite tumor type specific. For example, lung cancers show a strong enrichment of SBS4, a signature in which G to T mutations are most prevalent. SBS4 is typical for lung cancers in tobacco smokers (*22*). The patterns of these lung cancer G to T mutations are best correlated with mutations induced by polycyclic aromatic hydrocarbons, exemplified by benzo[*a*] pyrene (*22*–*26*).

G to T signatures have been extracted from a few other specific cancer types, in which they occur along with additional signatures. From the signature collection of the COSMIC database (*20*), SBS18 has been suggested to be potentially caused by ROS. SBS18 is most common in esophageal adenocarcinomas and in certain stomach and colorectal cancers. Esophageal adenocarcinomas have a clear preinflammatory condition, Barrett's esophagus. Rare colorectal tumors carry germline or somatic mutations in the oxidative damage repair enzyme MUTYH (*16*). These mutations, which show a high frequency of G to T events, have been characterized as SBS36 in the COSMIC database (*27*).

One important goal is to have a clear understanding of the extent and genomic features of the oxidative DNA damage that may cause the mutations. In most cases, DNA damage cannot be read directly by standard DNA polymerase–based sequencing methods. There are several methodologies available for mapping of oxidative DNA damage (*28*). Using immunoprecipitation of DNA with antibodies against 8-oxo-dG, followed by high-throughput sequencing, it is possible to map 8-oxo-dG in the genome with a resolution of 150 to 250 base pairs (bp) (*29*, *30*), which is analogous to resolution achieved by chromatin immunoprecipitation (ChIP) sequencing. Other affinity-based pulldown methods are 8-oxo-G-sequencing (OG-seq) (*31*) and apurinic site-sequencing (AP-seq) (*32*), which achieve similar levels of resolution.

However, it is most desirable to achieve single-base specificity for mapping of oxidized DNA bases. One method developed for this purpose is based on click chemistry (*33*). This approach involves removing the damage with a repair enzyme and filling the gap with an alkynylated deoxynucleoside, which can then be coupled to a code sequence oligonucleotide. The resulting triazole-linked DNA can be traversed and read by a DNA polymerase. This method has been used to map 8-oxo-dG in the yeast genome (*33*). Nick-sequencing (nick-seq) has been used for mapping oxidative DNA damage in *Escherichia coli* (*34*). 8-oxo-G can be further oxidized and chemically labeled with biotin, which produces a polymerase stop signal (*35*). However, to our knowledge, there is currently no straightforward enzyme-based method to map oxidized DNA bases genome-wide and at single-base resolution in mammalian cells.

We recently developed the circle-damage-sequencing (CD-seq) method for base level mapping of UV-induced DNA damage in human cells (*36*). This method relies on the efficient conversion of DNA damage into strand breaks and bidirectional sequencing from the break. Here, we have adopted CD-seq for mapping the sequence distribution of 8-oxo-G and of oxidized thymine in human cells. These damaged bases form with unique sequence specificities. The patterns of 8-oxo-dG resemble two mutational signatures found in cancer genomes.

RESULTS

Mapping of oxidized guanines at base resolution

We used hydrogen peroxide as the ROS with physiological relevance and incubated human dermal fibroblasts with this oxidizing agent at concentrations ranging from 5 to 25 mM for 30 min. We isolated DNA from the treated cells and initially verified the formation of oxidized guanines by cleavage of the high–molecular weight DNA with the Fpg DNA glycosylase. Fpg protein efficiently excises 8-oxo-G. It also releases ring-opened formamidopyrimidine (FAPY) DNA adducts (*37*–*39*), which may form under reducing conditions through one-electron reduction after free-radical reaction of guanine (*1*, *40*). After incubation of the DNA with Fpg protein, we produced DNA double-strand breaks at the incision sites using single-strand–specific S1 nuclease. We separated the treated DNA molecules on nondena-turing agarose gels (fig. S1). Increasing concentrations of H_2O_2 produced dose-dependently increasing levels of strand breakage, consistent with formation of 8-oxo-dG in the treated cells at frequencies of about one modification per 10 kb at the highest concentration used and less at the lower concentrations.

Having confirmed the formation of 8-oxo-dG, we selected the samples treated with 5 and 10 mM H_2O_2 and processed them for modified base mapping using CD-seq (Fig. 1A) (*36*). After DNA circularization, the 8-oxo-G bases were released with *E. coli* Fpg protein, which also has AP lyase activity. We used APE1 to remove the modified sugar residues at the 3′ ends near the single-strand breaks. Using S1 nuclease, we then produced ligatable DNA double-strand breaks at the ring-opened molecules. After adapter ligation, sequencing libraries were prepared and subjected to paired-end sequencing on Illumina flow cells (see Methods). In the CD-seq procedure, the aligned reads show a unique pattern in genome browser views; they appear as divergent reads (due to opening of the rings) with a single-nucleotide gap, whereby the gap position represents the base that was eliminated by the DNA glycosylase enzyme (Fig. 1, B to D). Considering all divergent gapped bases obtained, we observed that guanine was the base in the gaps at a frequency of over 70% of all bases. Adenine or thymine appeared in ~30% of the divergent reads. The presence of adenine may be caused by adenine oxidation or, alternatively, by the presence of abasic sites at adenine positions, which are cleaved by the AP lyase activity inherent to Fpg protein. However, in terms of the number of total divergent reads normalized to the total number of all reads obtained, our untreated control sample had 500 to 700 times fewer divergent reads with single-base gaps, suggesting that most of the damaged sequence positions we mapped were due to the treatment with H_2O_2 and were not due to a background of such modifications in cells or due to artifactual damage to the DNA during sample processing. The untreated samples contained 54% A or T and 46% G or C bases in the gap. The H_2O_2-treated cell samples had predominantly G or C in the gap at frequencies of 67 to 77.5%. One technical challenge with mapping oxidative DNA damage is its potential occurrence as background,

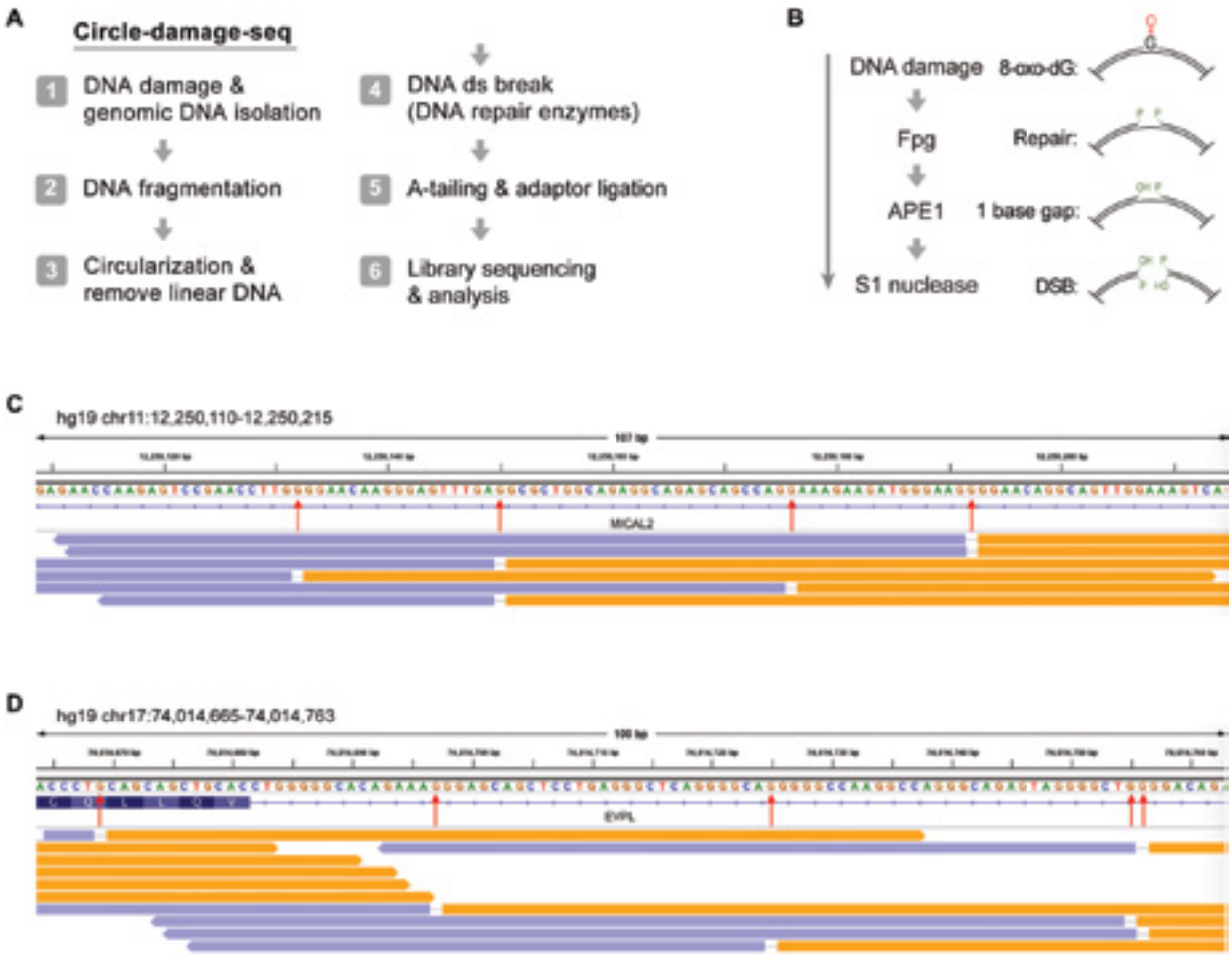


Fig. 1. Mapping of oxidized guanines by CD-seq. (**A**) Outline of the CD-seq method. (**B**) Outline of the enzymatic cleavage procedure to map 8-oxo-dG at single-base resolution. (**C**) Genome browser view of divergent DNA sequencing read pairs on chromosome 11 at the *MICAL2* gene. Orange and lavender segments represent the divergent read pairs. Red arrows indicate single-base gaps matching the excised DNA base. (**D**) Genome browser view of divergent DNA sequencing read pairs on chromosome 17 at the *EVPL* gene. ds, double strand; DSB, double-strand break.

which could be induced in part by the DNA preparation methods, most notably in the presence of phenol (*41*). Therefore, care needs to be taken to minimize background oxidation.

Distribution of oxidized guanines along genes

Using 5 million to 6 million divergent single-base gapped reads from cells treated with 5 or 10 mM H_2O_2 and much fewer reads from non-treated cells (n = ~12,000), the reads were mapped to the hg19 human genome. The oxidized bases were most strongly enriched in intergenic regions and in introns (Fig. 2A). Using gene-level (metagene) analysis, we profiled the oxidized base signals along all genes from −3 kb upstream of transcription start sites (TSSs) and then binned them according to gene length from the TSS to the transcription end sites (TESs) and continuing 3 kb downstream of the TES (Fig. 2B). The H_2O_2-treated samples showed similar profiles between the different concentrations, which were characterized by strong peaks near the TSS and a dip near the TES (Fig. 2, B to E). This distribution of signal roughly followed the GC-content of the human genome as plotted in Fig. 2F. Sequences near the TSS are generally GC rich and often include CpG islands, and sequences near the TES are AT rich due to the presence of polyadenylation signals and their surrounding sequences (*42*, *43*).

Relationship of guanine oxidation to gene expression and chromatin features

Using publicly available gene expression data for human fibroblasts, we examined whether 8-oxo-dG formation is correlated with gene expression levels. Figure S2 shows that there is no correlation between transcript levels (FPKM, fragments per kilo base per million mapped reads) and the extent of 8-oxo-dG formation in the region covering 2.5 kb upstream of the TSS.

Using chromatin state descriptions from the ENCODE (The Encyclopedia of DNA Elements) project (ChromHMM), we determined whether higher levels of 8-oxo-G bases accumulate in genomic regions that are associated with specific chromatin features (fig. S3). The highest levels of oxidized guanines were found in intragenic enhancer regions, in bivalent enhancers and TSS regions, and in genomic regions targeted by the Polycomb complex. Regions of heterochromatin and quiescent regions of the genome showed the lowest levels of 8-oxo-dG formation (fig. S3). These data are consistent with higher GC content of specific genomic segments being associated with higher levels of guanine oxidation, with limited contribution of gene expression state or specific chromatin features.

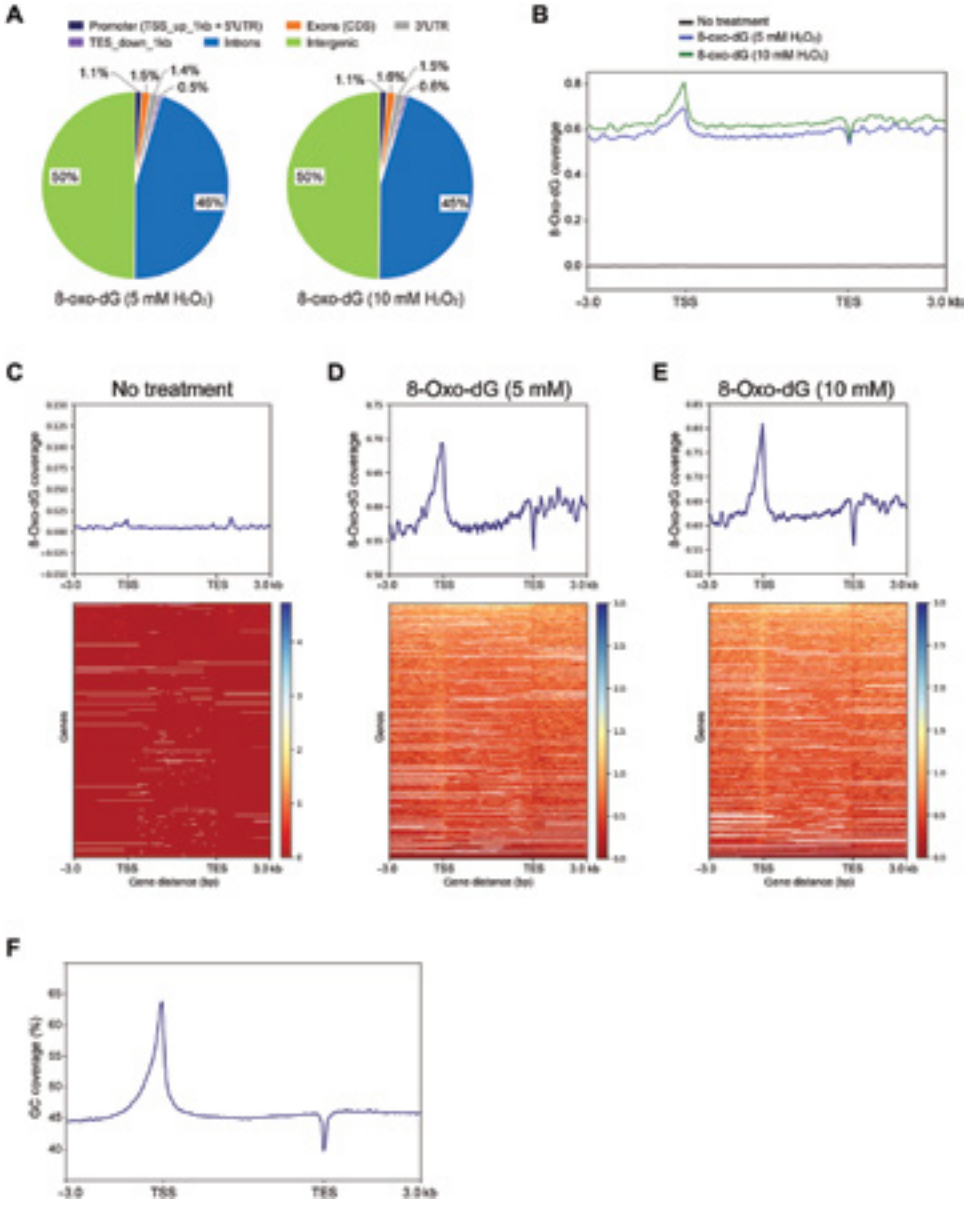


Fig. 2. Distribution of 8-oxo-dG along human genes. (**A**) Distribution of 8-oxo-dG in different compartments of the human genome. (**B**) Metagene profiles of 8-oxo-dG distribution along all genes of the hg19 human genome. The signal was plotted from 3 kb upstream of TSSs and then binned over the entire gene length, continuing 3 kb downstream of TESs. The *y*-axis scale is 0 to 0.8. (**C** to **E**) Heatmaps of 8-oxo-dG coverage are sorted from high (top, blue) to low (bottom, red). The signals were mapped and binned in 50-bp windows from 3 kb upstream of the TSS and then normalized relative to gene length over the gene bodies to the TES and 3 kb downstream of the TES. (C) No H_2O_2 treatment. (D) Treatment of cells with 5 mM H_2O_2. (E) Treatment of cells with 10 mM H_2O_2. (**F**) GC content along hg19 genes.

A previous report, using an immunoprecipitation technique, indicated that DNA in the nuclear periphery is more susceptible to H_2O_2-induced guanine oxidation than DNA in the center of the nucleus (*29*). The nuclear periphery–located DNA contains lamin-associated domains (LADs), which are regions of heterochromatin (*44*). We used published LAD data from an earlier study of human fibroblasts (*45*) and determined the distribution of 8-oxo-dG in LAD regions and in LAD-flanking regions, up to 100 kb from the border regions. 5-Oxo-dG was moderately depleted in LADs and enhanced at the border and flanking regions (fig. S4A). In addition, here, the distribution of the oxidized bases largely followed the GC content of the genome, which is lower in LADs (fig. S4B).

Distribution of oxidized guanines along specific DNA sequences

We next analyzed the trinucleotide sequence contexts of guanine oxidation in H_2O_2-treated human cells. Because the strands are symmetrical, for example, AGC corresponds to GCT on the opposite strand, we considered 48 combinations because we treated A or T in the gap separately. The middle base is displayed as either G, A, or T, and the 5′ and 3′ flanking bases can be any of the four

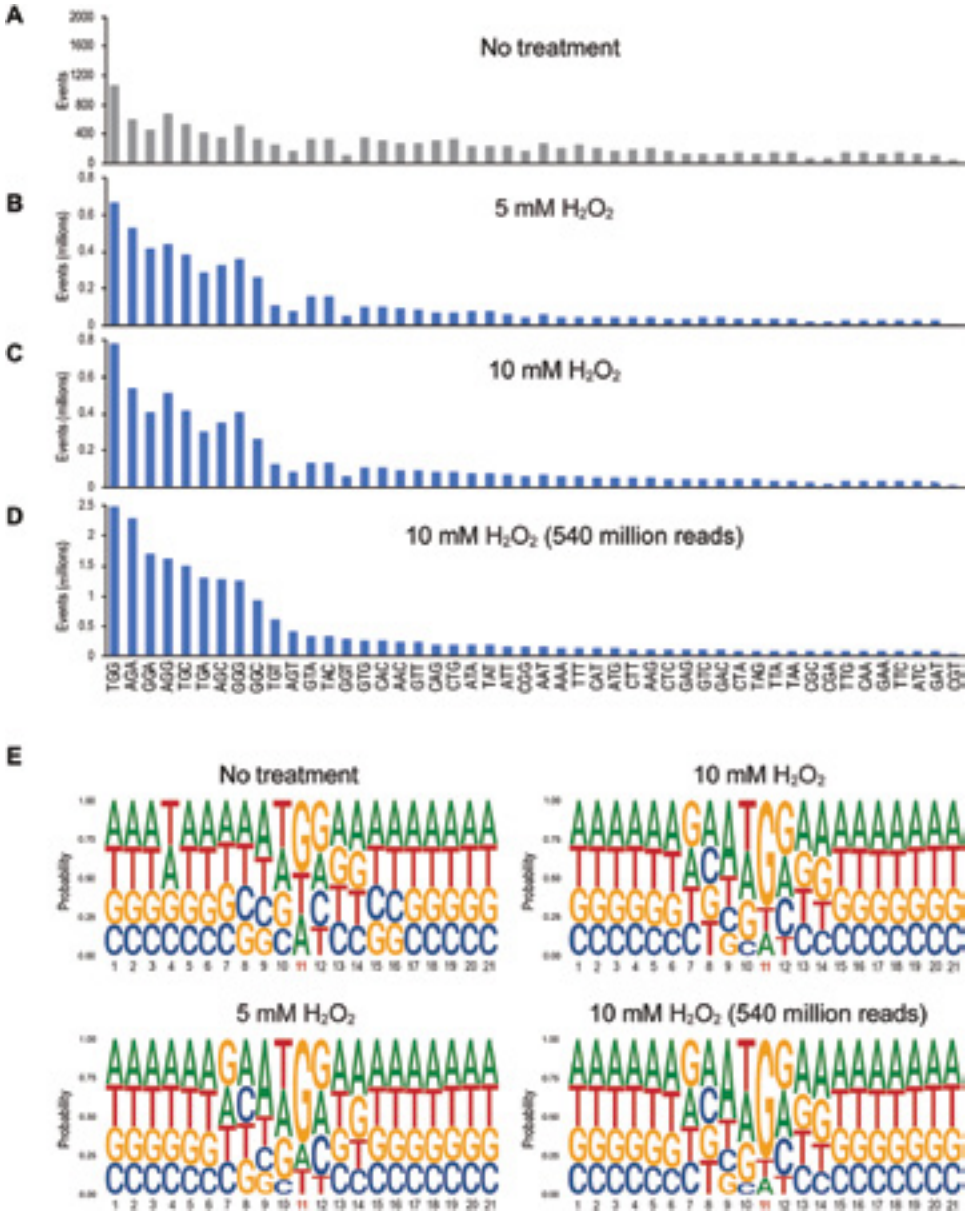


Fig. 3. Sequence context of 8-oxo-dG formation in human fibroblasts treated with H_2O_2. (**A**) Trinucleotide sequence context of read gaps in control cells (no treatment). (**B** to **D**) Trinucleotide sequence context of read gaps in cells treated with 5 mM (B) or 10 mM (C and D) H_2O_2. The read depth was 90 million reads in (C) or 540 million reads in (D). (**E**) Sequence context of guanine oxidation by using logo plot analysis. Position 11 represents the base in the divergent read gaps. At each base position, the height of each letter represents the relative frequency of that nucleic acid base.

bases, which leads to 48 combinations. For the nontreated control, the distribution of the different trinucleotides was relatively flat, with some enhancement at the sequences 5′TGG, 5′AGG, and 5′GGG (Fig. 3A). However, for the samples treated with 5 or 10 mM H_2O_2, trinucleotides with a central guanine were much enhanced (Fig. 3, B to D, consider y-axis scales compared to Fig. 3A). Here, we also included a sample derived from a 10 mM H2O2 treatment that was run on the sequencer at 540 million reads (Fig. 3D). The most highly damaged trinucleotide sequence always was 5′TGG. Additional major damage sites were found at 5′AGA, 5′AGG, and 5′GGA. Two neighboring purines with G as the damaged base as well as the sequence 5′TGC characterized the most frequent trinucleotides harboring oxidized guanine in the center. Expanding the analysis to a broader sequence context, we generated logo plots (Fig. 3E). These sequence plots show guanine as the preferred base in position 11 (the base in the gaps of divergent reads). The bases most preferred in the 5′ direction were T, A, and G. Cytosine was clearly underrepresented, largely because 5′CG sequences are depleted in mammalian genomes. The bases most preferred in the 3′ direction of the 8-oxo-G were G, A, and C, with T being strongly underrepresented. Outside of this trinucleotide context, no further sequence-specific enrichment of 8-oxo-G was apparent.

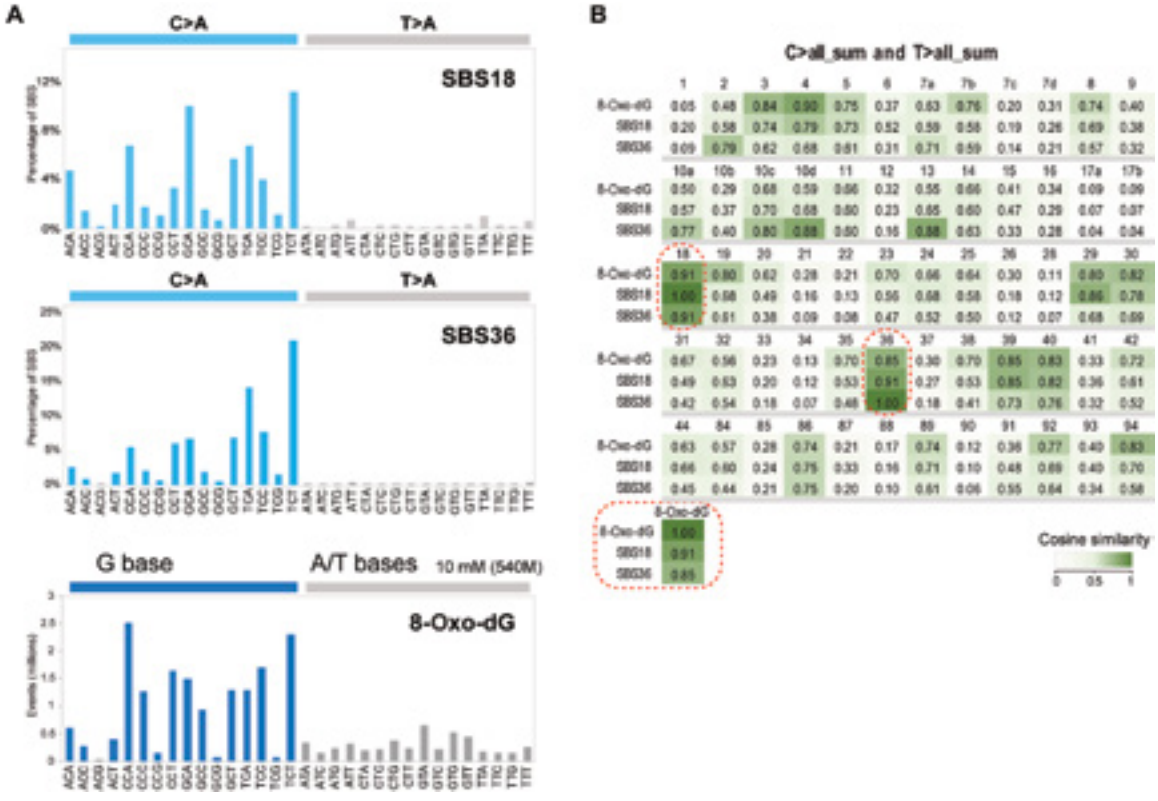


Fig. 4. 8-Oxo-dG signature and cancer mutational signatures. (**A**) COSMIC signatures SBS18 and SBS36 in comparison to the 8-oxo-dG signature. For the cancer signatures, only the C>A and T>A mutation windows are shown. The 8-oxo-dG DNA damage signature is displayed to reflect the style of cancer mutation signatures. (**B**) Heatmap showing the cosine similarity scores for 8-oxo-dG mapping at the trinucleotide level and the COSMIC mutational signatures found in the COSMIC v3.2 mutation database. We used the sum of all C to N and T to N mutation windows for comparisons with the 8-oxo-dG trinucleotide signature. We excluded COSMIC signatures that are thought to be sequencing artifacts. Darker green colors indicate higher similarity. Signatures SBS18 and SBS36 are indicated by dotted round rectangles.

It is of interest to compare this 8-oxo-dG DNA damage signature with the set of mutation signatures collected in the COSMIC cancer mutation database. We used cosine similarity analysis to establish the relationship between our 8-oxo-dG signature (established from 540 million reads) and the various COSMIC signatures. The highest similarities between the 8-oxo-dG signature and COSMIC SBS signatures were found for SBS4, SBS18, SBS36, and SBS39 (Fig. *4*). SBS4, which is enriched in G to T mutations, has been explained by tobacco smoke–associated polycyclic aromatic hydrocarbons (*25*). SBS39, a signature of unknown etiology, has predominantly G to C transversions, in addition to C to A mutations, and would therefore not qualify for a specific 8-oxo-dG signature. Those signatures that are dominated by C to A (or G to T) transversions are of greater relevance.

SBS18 has been proposed to be linked to ROS-induced DNA damage (COSMIC database). SBS36 is thought to be due to deficiency of the MUTYH DNA glycosylase known to excise adenines from 8-oxo-G/A mispairs, is enriched in colorectal tumors from patients with *MUTYH* mutations, and therefore should reflect a signature from oxidative DNA damage (*27*, *46*). A similar signature of mutations has been found in *Mth1*/*Ogg1*/*Mutyh* triple knockout mice (*47*, *48*). The cosine similarity between the 8-oxo-dG signature and SBS18 was highest at 0.91, and the cosine similarity between 8-oxo-dG and SBS36 was 0.85 (Fig. 4B). Our analysis suggests that the trinucleotide patterns we established for oxidized guanines are concordant with two G/C to T/A mutational signatures enriched in human cancers.

Analysis of oxidized thymines in permanganate-treated cells

Potassium permanganate oxidizes thymine and produces thymine glycol as the main DNA oxidation product (*11*). The sequence specificity and genomic preferences for thymine oxidation have remained unknown due to a lack of available methodology.

To analyze permanganate-oxidized thymines, we treated human U2OS cells with potassium permanganate and isolated the DNA. To test the global extent of thymine oxidation in the treated cells, we initially used agarose gel electrophoresis after cleavage of the DNA with two DNA glycosylases, Endo III and NEIL1. These enzymes remove oxidized DNA bases from DNA (*49*, *50*). Whereas Endo III prefers to excise oxidized bases from double-stranded DNA, NEIL1 preferentially operates on single-stranded regions (*51*), although there likely is overlap in their substrate preference depending on enzyme concentrations and incubation conditions. The enzyme Endo IV was used as an AP endonuclease to remove ligation-blocking lesions from the 3′ ends. Double-strand cleavage was achieved with S1 nuclease digestion. The data show that the combined enzyme treatment produces double-strand breaks in a $KMnO_4$ concentration–dependent manner (fig. S5) and that the treatment with 10 mM $KMnO_4$ creates approximately one lesion per 3 kb of DNA.

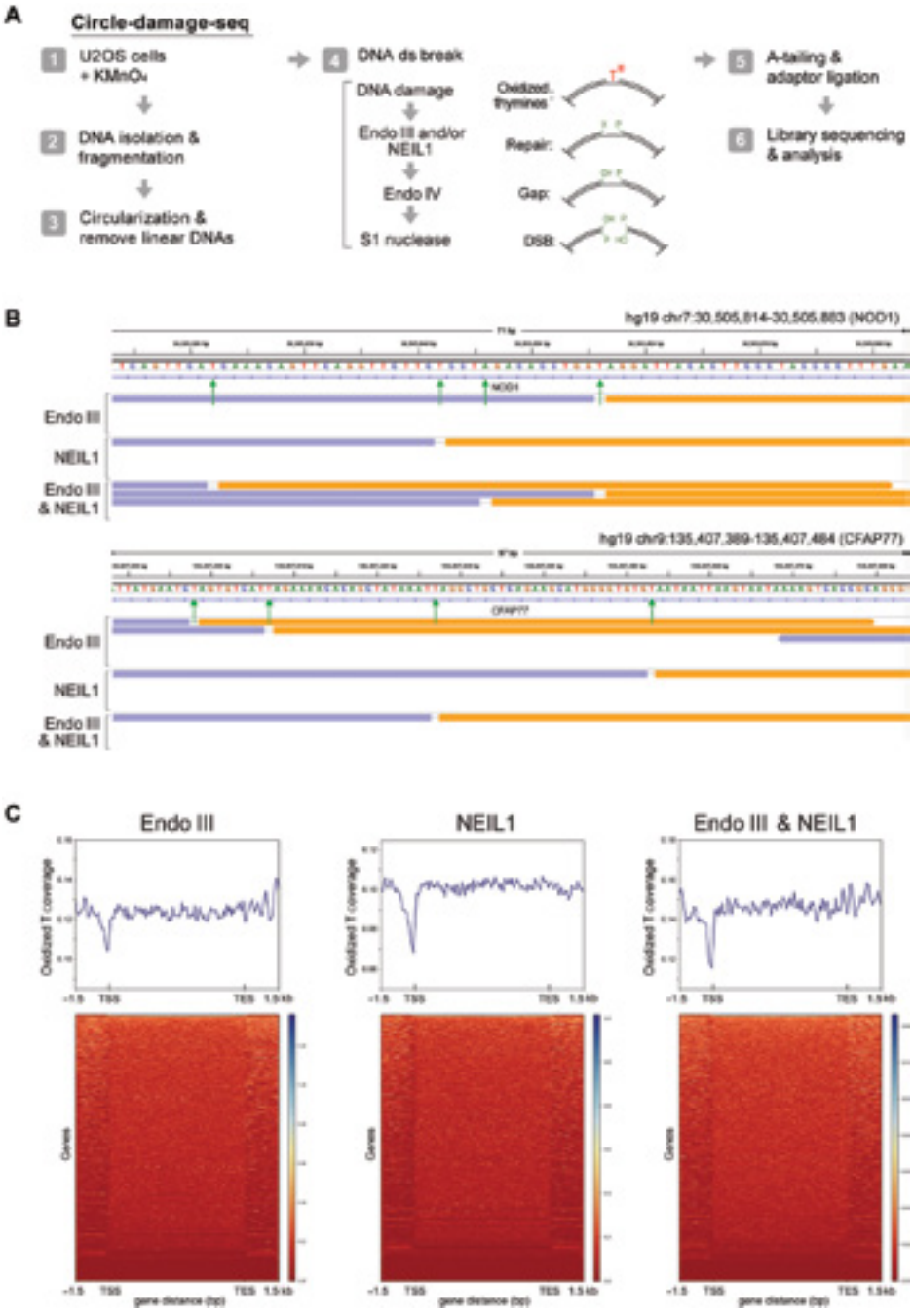


Fig. 5. Mapping of oxidized thymines. (**A**) Outline of the CD-seq and enzymatic cleavage methods used to map oxidized thymines at single-base resolution. (**B**) Base resolution view of oxidized thymines along the human *NOD1* and *CFAP77* genes. Orange and lavender segments represent the divergent read pairs. Green arrows indicate single-base gaps matching the excised oxidized DNA base. (**C**) Heatmaps and metagene profiles of permanganate-oxidized thymines along human genes. DNA from untreated cells produced very few divergent read pairs, and the data could not be used in similar displays. Different cleavage enzymes, Endo III or NEIL1, or a combination of the two was used.

Following these initial experiments, we used NEIL1 or Endo III in the CD-seq procedure (Fig. 5A). We also included samples in which both enzymes were used in combination. Figure 5B shows high-resolution genome browser views of the amplified libraries. As expected, we almost exclusively observed thymines in the single-base gaps, consistent with the properties of permanganate oxidation and DNA glycosylase specificity.

Distribution of oxidized thymines along genes

We created several million divergent reads with single-base gaps and mapped them to the hg19 genome. We analyzed the damage detected with the different enzyme combinations along genes from –3 kb upstream of TSSs and then binned according to gene length from the TSS to the TES, and 3 kb downstream of the TES (Fig. 5). These profiles show a pronounced reduction of oxidized thymine

near TSS regions and a relatively flat profile along other parts of the genes. It is likely that the reduced levels of oxidized Ts near the TSS are a consequence of the GC richness of these sequences.

The levels of oxidized thymines in regions representing upstream gene promoters did not correlate with gene expression levels (fig. S6). We then derived ChromHMM maps for human U2OS cells using published ChIP-sequencing (ChIP-seq) data for histone modifications (fig. S7). We found the strongest enrichment of oxidized thymines at zinc finger genes and at repeats (fig. S7A). The enrichment of thymine glycol in zinc finger genes may be explained, at least in part, by the relative AT richness (GC poorness) of these genes relative to other hg19 genes (fig. S7B). However, with regard to LADs, the distribution of oxidized thymines was uniform along those regions and flanking sequences and was not negatively correlated with GC content of the regions (fig. S8).

Distribution of oxidized thymines along specific DNA sequences

We next analyzed the trinucleotide sequence contexts of thymine oxidation damage in permanganate-treated cells. Because the strands are symmetrical, for example, ATC corresponds to GAT on the opposite strand, we considered 48 combinations. The middle base is displayed as either T, C, or G, and the 5′ and 3′ flanking bases can be any of the four bases, which creates 48 combinations. For

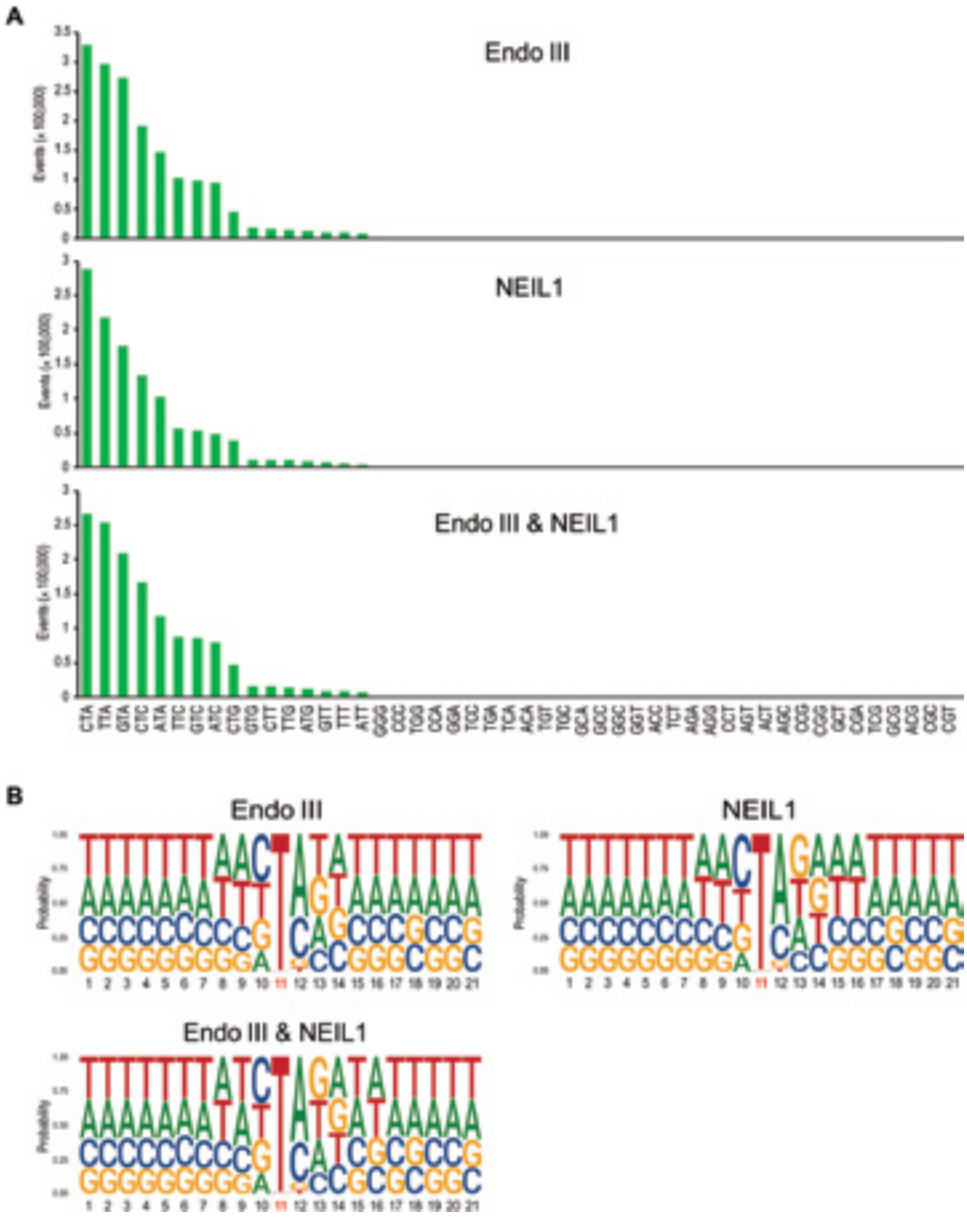


Fig. 6. Sequence-level distribution of oxidized thymines in permanganate-treated human cells. (**A**) Trinucleotide sequence specificity. Forty-eight trinucleotide contexts were analyzed, but the gapped reads occurred chiefly at trinucleotides with central thymine. Different enzymes (Endo III and NEIL1 in combination and alone) were used to excise the oxidized bases. (**B**) Sequence context of thymine oxidation by permanganate using logo plot analysis. Position 11 represents the permanganate-oxidized thymine base. At each base position, the height of each letter represents the relative frequency of that nucleic acid base. Note the enrichment of A or C 3′ to the thymine.

non-treated cells, we obtained only 392 divergent reads with single-base gaps. One reason for this very small number of background reads is the low frequency of oxidized pyrimidines in untreated cells. The number of reads was too low to establish a reliable trinucleotide distribution list or gene profiles for nontreated cells. The different enzyme treatments with Endo III, NEIL1, or the combination of the two gave very similar trinucleotide sequence patterns. The preferred sequence contexts of oxidized thymines were 5′CTA and 5′TTA, followed by 5′GTA, 5′CTC, and 5′ATA (Fig. 6A).

A sequence logo plot analysis (Fig. 6B) revealed T as the overwhelming base in position 11, the gapped position of divergent reads. The position 5′ to the damaged T showed a minor preference for pyrimidines. However, the position 3′ to the oxidized T showed interesting characteristics, the common presence of A followed by C and much fewer occurrences of G or T. The data reveal a heretofore unknown DNA sequence specificity of permanganate-induced thymine oxidation.

DISCUSSION

We applied the CD-seq method for comprehensive mapping of two types of oxidative DNA damage. CD-seq is applicable for analysis of any type of DNA damage or DNA modifications for which a specific cleavage enzyme is available. In our study, we used DNA glycosylases of the base excision repair pathway. These enzymes may have a narrow or broader selectivity for the damaged bases. In this case, Fpg protein is known to excise 8-oxo-G from DNA but also acts on other oxidized bases, preferentially purines, such as FAPY-G or oxidized adenines. The DNA damaging agent we used, H_2O_2, is known to induce 8-oxo-G as a major modified DNA base. Therefore, the signals we mapped are expected to be mostly 8-oxo-G. For permanganate-induced DNA damage, we used *E. coli* Endo III, which preferentially recognizes oxidized pyrimidines. We also used the NEIL1 protein, which has a broader specificity and can remove oxidized purines and oxidized pyrimidines. The DNA damaging agent $KMnO_4$ oxidizes thymine, preferentially leading to the formation of thymine glycol (*10*, *11*). Notably, the sequence patterns we obtained with Endo III and NEIL1 cleavage were almost identical and consisted of almost 100% thymine, which suggests that the procedure mapped thymine oxidation products. Our work revealed a previously unrecognized sequence specificity of permanganate oxidation, 5′TA/C.

Oxidized guanines obtained after treatment of cells with H_2O_2 formed with a defined sequence pattern in human cells. This oxidizing agent is present endogenously in many cell types and tissues. H_2O_2 is generated as a product of white blood cells that are active during immune responses and under inflammatory conditions. In the presence of transition metal ions, hydroxyl radical and other oxidizing short-lived intermediates are formed, which damage guanine preferentially. We found that guanine oxidation by H_2O_2 is chiefly related to the GC content of the genome with GC-rich regions, such as TSSs, acquiring the highest levels of 8-oxo-dG. This result is partially consistent with previous results using DNA immunoprecipitation, in which the authors observed the highest levels of 8-oxo-dG just downstream of the TSS (*52*). However, Poetsch *et al.* (*32*) reported a substantially decreased AP and 8-oxo-dG signal using their AP-seq method near TSS regions. This contrasts with our observations, but neither one of these two earlier publications used H_2O_2 for treatment of the cells. The base oxidation patterns were not noticeably dependent on gene expression state and chromatin environments except that such environments have differential GC content.

Oxidative DNA damage has long been suspected to have a causative role in cancer, aging, and other diseases such as neurodegeneration. However, definitive proof of this connection has been difficult to obtain. Regimens of diets rich in antioxidants are thought to prevent cancer and promote longevity. Studies in mouse models defective in DNA repair pathways that deal with oxidative DNA damage are perhaps the best available evidence to support the oxidative damage and cancer relationship. These studies, for example, with *Ogg1* knockout mice, often do not lead to a major cancer phenotype, presumably because of effective antioxidant defense systems and redundant DNA repair capacities (*53*, *54*). However, mice double-deficient in OGG1, which removes 8-oxo-G from 8-oxo-G/C base pairs, and deficient in MUTYH, which removes adenine from 8-oxo-G/A mispairs, are clearly cancer prone (*18*).

For human cancers, a direct relationship between oxidative DNA damage and cancer has been even more difficult to obtain. Although inflammation is a cancer-predisposing condition, which has been suspected to contribute to a large fraction of human tumors, the inflammatory process induces many other changes in addition to direct DNA damage. For example, inflammation is contributing to altered signaling pathways, altered immune responses, changes in gene expression, and epigenetic changes such as DNA hypermethylation (*55*).

There has been great technical difficulty in precisely measuring endogenous oxidative DNA damage in tissues. The methods used are often confounded by artifactual damage introduced during DNA isolation or sample processing (*41*, *56*, *57*). Therefore, it has been challenging to determine DNA damage in human tissues including precancerous lesions. With further improvements and effective precautions to prevent in vitro damage to DNA (such as inclusion of antioxidants during DNA isolation), the CD-seq method may eventually achieve high enough sensitivity to measure oxidized DNA bases in human tissue specimens. In a previous study, Kucab *et al.* (*25*) used cell cloning and mutation sequencing as an end point to measure mutagenesis induced by H_2O_2 but did not observe clear mutation patterns. This approach requires DNA replication, which necessitates using a rather low concentration of hydrogen peroxide. To produce sufficient DNA damage, in our hands required millimolar concentrations of H_2O_2, which may not permit DNA replication to occur due to cell cycle arrest.

Human cancer genome sequencing efforts have provided catalogs of cancer-specific mutational patterns for dozens of human tumor types. Several of these mutational patterns give an indication for what the premutagenic DNA damaging processes might be. For example, sunlight-induced skin cancers carry a clear mutation signature (referred to as SBS7a and SBS7b in the current COSMIC database). This signature consists chiefly of C to T mutations at dipyrimidine sequences, for example, at 5′TC. Using CD-seq, we recently showed that this signature is highly similar to the distribution of UVB irradiation–induced cyclobutane pyrimidine dimers that have undergone cytosine deamination (*36*). For oxidative DNA damage, the signatures should be dominated by G to T (or C to A) transversions, which are known mutations induced by 8-oxo-dG (*13*, *14*). There are several COSMIC signatures that have these required features. One is SBS4, which is smoking associated and has been linked to tobacco-related polycyclic aromatic hydrocarbons (*22*, *25*, *58*). SBS10c, SBS10d, SBS14, and SBS20, all enriched in G to T mutations, have been linked to DNA polymerase mutations. SBS24 is linked to aflatoxin exposure, and SBS29 is found in tobacco chewers.

Another G to T mutation signature is SBS18, which is of unproven etiology. When we compared our 8-oxo-dG signature with SBS18, we obtained a strong cosine similarity value of 0.91, which was higher than that with any other signature (Fig. *4*). We point out that mutational signatures are not solely the product of a one-to-one relationship between DNA damage and mutations. The mutations are shaped by additional factors that are currently unknown such as a potential sequence-specific repair of 8-oxo-G (*46*) or a sequence-dependent bypass of the oxidized guanines by DNA polymerases. Nonetheless, SBS18 is so similar to the 8-oxo-dG signature that they are likely directly related. Because repair of 8-oxo-G follows rapid kinetics with most damage removed within minutes (*59*, *60*), our measurements most likely reflect an equilibrium between ongoing H_2O_2-induced damage formation and repair of this damage in cells. We propose that this equilibrium is perhaps the best way of reflecting mutational patterns because it is the equilibrium (unrepaired) DNA damage that will translate into mutations. Knockout of OGG1 in human untreated induced pluripotent stem cells also produced mutations resembling SBS18, albeit at lower similarity levels (*46*). It is worth noting that SBS18 is most enriched and prevalent in gastrointestinal cancers that have a strong inflammatory component, such as esophageal adenocarcinoma (linked to Barret's esophagus), gastric adenocarcinoma (linked to gastritis), and a few colorectal cancers (fig. S9). Colonic crypts from patients with inflammatory bowel disease show an increase of SBS18 mutations (*61*). Inflammation is strongly associated with ROS (*5*, *6*).

SBS36 is also defined by G to T/C to A transversions (Fig. 4A) and is quantitatively similar to SBS18 (cosine = 0.91 between the two signatures). This signature is found in colorectal cancers of patients that have biallelic germline mutations or somatic mutations in the *MUTYH* gene (*27*). Lack of MUTYH activity will eliminate an important correction pathway that would otherwise remove mispaired adenine opposite 8-oxo-G. When this adenine is not removed, mutations at 8-oxo-dG will be created and fixed more readily. For these reasons, SBS36 will be a good indicator of 8-oxo-dG–induced mutations. Comparison of our 8-oxo-dG trinucleotide distribution data with SBS36 indicated a cosine similarity value of 0.85, which is among the highest observed for all available COSMIC signatures. Another type of tumor in which a ROS-mediated mutagenesis pathway seems to be at work is neuroblastoma with copy number loss of *OGG1* or *MUTYH* as recently reported (*62*).

In conclusion, we present a method based on CD-seq, which can be used to map various types of oxidatively damaged DNA bases at single-base resolution in the human genome. We define a sequence-specific pattern for oxidized guanines induced by H_2O_2 and of oxidized thymine induced by permanganate. The 8-oxo-dG trinucleotide distribution pattern closely resembled cancer mutation signatures SBS18 and SBS36, suggesting the involvement of guanine oxidation in specific types of human cancers of the gastrointestinal tract.

METHODS

Cell culture

Human male skin fibroblast cells [human dermal fibroblast (HDF); American Type Culture Collection (ATCC), PCS-201-012] with no genetic modification and human osteosarcoma cells (U2OS; ATCC, HTB-96) were cultured in the Fibroblast Growth Kit (ATCC, catalog no. PCS-201-041) supplemented with 2% fetal bovine serum (FBS) and in Dulbecco's modified Eagle's medium/high glucose supplemented with 10% FBS, respectively, at 37°C in a 5% CO_2 standard incubator.

Oxidative DNA damage and genomic DNA isolation

To induce 8-oxo-dGs, HDF cells at 80 to 90% confluence on 10-cm culture plates were washed with 1× phosphate-buffered saline (PBS) and treated in PBS with H_2O_2 solution diluted to final concentrations of 5, 10, and 25 mM for 30 min at 37°C in a 5% CO_2 incubator. After treatment, the cells were immediately trypsinized and pelleted, and then genomic DNA was isolated using the Quick-DNA Miniprep Plus Kit (Zymo Research, Irvine, CA) according to the manufacturer's instruction manual. To evaluate 8-oxo-dG damage in the isolated genomic DNAs, the DNAs were treated with Fpg and APE1 [New England Biolabs (NEB), Ipswich, MA] followed by S1 nuclease (Thermo Fisher Scientific) treatment, and then damage-specific DNA cleavage events were observed on 1% agarose gels.

To induce oxidized thymines, 1×10^7 U2OS cells were washed with 1× PBS and treated with 10 or 40 mM $KMnO_4$ for 70 s at 37°C in a solution with the following components: 15 mM tris-HCl (pH 7.5), 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, and 300 mM sucrose. The reaction was quenched by adding β-mercaptoethanol to a final concentration of 700 mM and EDTA to a final concentration of 50 mM. SDS was added to 1%, and the DNA was purified by proteinase K digestion, phenol-chloroform extraction, and ethanol precipitation. The quenching reaction and the proteinase K digestion were carried out at 37°C for 30 min and 4 hours, respectively. DNA base damage was evaluated by treatment with Endo III (NEB) and/or NEIL1 (OriGene, Rockville, MD) DNA glycosylases, followed by Endo IV and S1 treatment, and then observed on a 1% agarose gel.

CD-seq mapping of 8-oxo-dGs

CD-seq library preparations were performed as previously described (*36*) with slight modifications to generate damage-specific DNA double-strand breaks. Briefly, 2 μg of the genomic DNAs prepared from H_2O_2-treated cells was sheared to an average length of 300 bp by sonication with a Covaris E220 sonicator (Covaris, Woburn, MA) and end-repaired to prepare blunt-ended DNA with ribonuclease H, T4 DNA polymerase, and T4 polynucleotide kinase (NEB). To circularize DNA fragments, 1 μg of the blunt-ended DNA was incubated with T4 DNA ligase (NEB) in a final reaction volume of 200 μl by overnight incubation at 16°C. Then, to remove noncircularized linear DNA from the circularized DNA pool, the ligase-treated DNA was incubated with Plasmid-Safe ATP-Dependent DNase (Lucigen, Middleton, WI).

To generate DNA double-strand breaks at 8-oxo-dG sites, the circularized DNA was incubated in 1× NEBuffer 1 reaction buffer with 1 μl (8 U/μl) of Fpg protein (NEB) in a final volume of 40 ml for 1 hour at 37°C. The DNA was cleaned up with 72 μl (1.8×) of AMPure XP beads (Beckman Coulter, Indianapolis, IN) and eluted in 35 μl of 10 mM tris-HCl (pH 8.0). To remove the modified sugar residues at the 3′ ends near the single-strand breaks, the DNA from above was incubated in 1× NEBuffer 4 reaction buffer with 1 μl (10 U/μl) of APE1 (NEB) for 20 min at 37°C. The DNA was cleaned up and cleaved on the opposite strand of the nicked positions using 1 μl (5 U/μl) of single strand–specific S1 nuclease (Thermo Fisher Scientific) for 4 min at room temperature. Then, the reaction was stopped by adding 2 μl of 0.5 M EDTA and 1 μl of 1 M tris-HCl (pH 8.0) to the reaction mixture, and DNA was further incubated for 10 min at 70°C. The DNA sample was cleaned up with 72 μl (1.8×) of AMPure XP beads and eluted in 48 μl of 10 mM tris-HCl (pH 8.0).

The double-strand cleaved DNA was A-tailed with Klenow fragment exo- (NEB) and subsequently ligated with 50 nM (final concentration) of T-overhang NEBNext hairpin adaptors (NEB): 5′-phos-GATCGGAAGAGCACACGTCTGAACTCCAGTCdUACACTCTTTCCTACACGACGCTCTTCCGATC*T-3′ and with 50 nM of C-overhang adaptors synthesized by Integrated DNA Technologies (Coralville, IA): 5′-phos-GATCGGAAGAGCACACGTCTGAACTCCAGTCdUACACTCTTTCCTACACGACGCTCTTCCGATC*C-3′ (*; phosphorothioate linkage), and this reaction was followed by incubation with USER enzyme (NEB). The polymerase chain reaction (PCR)–amplified CD-seq library was sequenced as 150-bp paired-end sequencing runs on an Illumina Hiseq2500 platform to obtain between 100 million and 540 million reads.

CD-seq mapping of oxidized Ts

To generate double-strand breaks at oxidized thymine sites, the circularized DNAs prepared from $KMnO_4$-treated cells were initially incubated in 1× reaction buffer [20 mM tris-HCl (pH 8.0), 1 mM EDTA, 50 mM NaCl, and bovine serum albumin (0.1 mg/ml)] with 0.5 μl (10 U/μl) of Endo III (NEB) and/or 1 μl (0.24 μg/μl) of NEIL1 (OriGene) in a final volume of 20 μl for 30 min at 37°C. Then, the DNAs were cleaned up and treated with Endo IV (NEB) to remove ligation-blocking lesions from the 3′ ends, followed by S1 nuclease digestion to achieve double-strand cleavage. The cleaved DNAs were subjected to A-tailing, and this was followed by adaptor ligation with 50 nM (final concentration) of T-overhang NEBNext hairpin adaptors (NEB). The adaptor-ligated DNAs were treated with USER enzyme and amplified by PCR. The CD-seq library was sequenced as 150-bp paired-end reads on an Illumina Hiseq2500 platform to obtain 50 million reads.

Data analysis

Identification of oxidized DNA bases

To identify the positions of oxidized DNA bases, fastq files containing the CD-seq reads were processed using the pipeline as previously described (*36*). Briefly, all reads were trimmed from the 3′ end to 2 × 80 nucleotide length after removal of sequencing adapters and low-quality reads (phred < *20*) using Trim_Galore (https://bioinformatics.babraham.ac.uk/projects/trim_galore/). BWA-MEM (version 0.7.17) was then used to align the reads to the human reference genome (hg19) and to remove duplicate alignments with the "removeDups" SAMBLASTER (version 0.1.26) option. Only divergently aligned pairs of reads with a single-nucleotide gap between reads were retained by selecting SAM records with TLEN of 3 and saved as bam files for downstream analyses. TLEN = 3 represents divergently aligned reads with one nucleotide between mates. In addition, chromosome name, start position, stop position, and inferred strand (+/−) for all damaged base loci were written to a BED file as processed data, and this information has been deposited into the Gene Expression Omnibus (GEO) database under accession number GSE184820.

Visualization of read alignments using the IGV viewer

To view aligned reads on the Integrative Genomics Viewer (IGV) downloaded from the Broad Institute (http://software.broadinstitute.org/software/igv/), the TLEN = 3 filtered bam files were sorted by genomic coordinates and indexed using SAMtools (version 1.11). The bam files were loaded on IGV viewer (version 2.8.9), and the tracks were displayed across the hg19 human genome with a "view as pairs" option to show pairs together with a line joining the ends.

Metagene profiles and genomic region analysis

To analyze the distribution of oxidized DNA damage along the hg19 human reference genes, the TLEN = 3 filtered bam files were processed on the Galaxy web platform (https://usegalaxy.org/). The bam files were converted to bigwig files using bamCoverage at 20-bp resolution, and computeMatrix and plotHeatmap programs in deepTools were then used to assess the oxidized bases over genic regions that included 3 kb upstream of the TSS and 3 kb downstream of the TES with average coverage calculated in nonoverlapping 50-nucleotide bins and normalized to gene length. For analysis of GC content of the hg19 human genome, we downloaded the hg19. gc5Base.wigVarStep.gz file from the UCSC Genome Browser database and converted it to bigwig file format. Then, the mean GC content distribution was analyzed using the same process as for metagene profiles above. The range of GC content was scaled to 35 to 70% GC content. The TLEN = 3 filtered bam files were also processed using the "Read Distribution" program in RseQC package provided on the Galaxy web platform to analyze the distribution of oxidized DNA bases over genomic features of the hg19 human reference genes.

Trinucleotide sequence profiles and logo plots

Sequences for the trinucleotides representing the 5′ and 3′ flanking bases and the gapped base in the middle were identified and counted using BEDtools (*63*). For 8-oxo-dGs, when there is a cytosine as the gapped nucleotide, the damage would have occurred on the (−) strand; thus, reverse complement trinucleotides with G or C in the middle were combined. For oxidized thymine, an adenine as the gapped nucleotide would indicate that the damage would have occurred on the (−) strand; thus, reverse complement trinucleotide sequences with T or A in the middle were combined, and the 48 possible trinucleotide combinations were used for trinucleotide sequence profiles. Sequence logo plots for nucleotide frequencies surrounding damaged bases (−10 to +10 positions relative to the oxidized base) were drawn using the ggseqlogo package in R (*64*).

Cosine similarity analysis

SBS mutational signatures with trinucleotide frequencies (version 3.2) were obtained in numerical form from the COSMIC database (http://cancer.sanger.ac.uk/cosmic/signatures/SBS/). We excluded 18 of the COSMIC signatures that are thought to be sequencing artifacts. To directly compare the two datasets, cosine similarity analyses were performed using the R (v. 4.0.2) package "coop" (https://R-project.org/). Because the mutated base of SBS signatures is represented by the pyrimidine of the base pair in the middle position, reverse complement trinucleotide sequences from CD-seq were combined and represented by the pyrimidine in the middle base. In addition, we used the sum of all C to N and T to N mutation windows for comparisons with the CD-seq trinucleotide sequence signature.

Chromatin state analysis

To determine the relative enrichment of oxidative DNA damage along defined chromatin states, ChromHMM (*65*) was used to perform genome annotations and to compute the fold enrichment scores of each state for oxidized damage mapping data. Genomic segmentation files for the chromatin states (core 18-state model) based on epigenomic data for human dermis fibroblasts were obtained from the Roadmap Epigenomics Project (http://roadmapepigenomics.org) (ENCODE accession no. ENCSR071BVW), and the scores showing the relative enrichment of 8-oxo-dG were calculated using the "OverlapEnrichment" option in ChromHMM with default parameters.

To establish ("learn") the chromatin state model for U2OS cells, the "LearnModel" option in ChromHMM with 200-bp bin size

was used using histone modification ChIP-seq data obtained from GEO database accession no. GSE141081 for H3K4me1, H3K4me3, and H3K27ac, and from accession no. GSE31755 for H3K9me3 and H3K36me3. After annotation of 14 chromatin states using the state emissions data and the TSS neighborhood data generated by the ChromHMM LearnModel, the fold enrichment scores for oxidized thymine along defined chromatin states were obtained as described above.

Lamin-associated domains

To determine the distribution of oxidized bases over LAD regions, previously defined LAD data (n = 1302 LADs) for human fibroblasts were obtained from published sources (*45*) and U2OS LAD data (n = 1046 LADs) were downloaded from the GEO database (GES87831) (*66*). The ComputeMatrix program in deepTools was used to calculate the coverages of oxidized DNA bases over LADs and 100-kb regions flanking LADs in nonoverlapping 1-kb bins. The plotProfile program in deepTools was used to plot the average coverages computed by computeMatrix.

Distribution of oxidized bases and gene expression profiles

Transcript levels (FPKM, fragments per kilo base per million mapped reads) determined by RNA-seq for human dermis fibroblasts and for U2OS cells were obtained from the GEO database accession numbers GSE78610 and GES139190, respectively. The transcript data were converted to bed file format and sorted in descending order with FPKM value for downstream analysis. The datasets were grouped into detectable (FPKM > *0*) and nondetectable (FPKM = *0*) transcripts. To determine the coverages of oxidized bases in extended promoter regions (within 2.5 kb upstream of the TSS) of detectable transcripts, the ComputeMatrix program in deepTools was used to obtain matrices containing scores for coverage of oxidized bases over promoter regions of each transcript. Scatter plots showing the correlation between oxidized DNA bases and gene expression levels were drawn using scores computed by computeMatrix and corresponding FPKM values.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abn3815

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. S. Boiteux, F. Coste, B. Castaing, Repair of 8-oxo-**7**,8-dihydroguanine in prokaryotic and eukaryotic cells: Properties and biological roles of the Fpg and OGG1 DNA N- glycosylases. *Free Radic. Biol. Med.* **107**, 179–201 (2017).
2. J. Cadet, K. J. A. Davies, M. H. Medeiros, P. Di Mascio, J. R. Wagner, Formation and repair of oxidatively generated damage in cellular DNA. *Free Radic. Biol. Med.* **107**, 13–34 (2017).
3. M. D. Evans, M. Dizdaroglu, M. S. Cooke, Oxidative DNA damage and disease: Induction, repair and significance. *Mutat. Res.* **567**, 1–61 (2004).
4. M. Giorgio, G. I. Dellino, V. Gambino, N. Roda, P. G. Pelicci, On the epigenetic role of guanosine oxidation. *Redox Biol.* **29**, 101398 (2020).
5. Y. Yu, Y. Cui, L. J. Niedernhofer, Y. Wang, Occurrence, biological consequences, and human health relevance of oxidative stress-induced DNA damage. *Chem. Res. Toxicol.* **29**, 2008–2039 (2016).
6. P. Lonkar, P. C. Dedon, Reactive species and DNA damage in chronic inflammation: Reconciling chemical mechanisms and biological fates. *Int. J. Cancer* **128**, 1999–2009 (2011).
7. J. M. Flynn, S. Melov, SOD2 in mitochondrial dysfunction and neurodegeneration. *Free Radic. Biol. Med.* **62**, 4–12 (2013).
8. J. C. Niles, J. S. Wishnok, S. R. Tannenbaum, Peroxynitrite-induced oxidation and nitration products of guanine and 8-oxoguanine: Structures and mechanisms of product formation. *Nitric Oxide* **14**, 109–121 (2006).
9. H. Maki, M. Sekiguchi, MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. *Nature* **355**, 273–275 (1992).
10. K. Frenkel, M. S. Goldstein, N. J. Duker, G. W. Teebor, Identification of the cis-thymine glycol moiety in oxidized deoxyribonucleic acid. *Biochemistry* **20**, 750–754 (1981).
11. P. Rouet, J. M. Essigmann, Possible role for thymine glycol in the selective inhibition of DNA synthesis on oxidized DNA templates. *Cancer Res.* **45**, 6113–6118 (1985).
12. S. S. David, V. L. O'Shea, S. Kundu, Base-excision repair of oxidative DNA damage. *Nature* **447**, 941–950 (2007).
13. S. Shibutani, M. Takeshita, A. P. Grollman, Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxo-dG. *Nature* **349**, 431–434 (1991).
14. M. L. Wood, M. Dizdaroglu, E. Gajewski, J. M. Essigmann, Mechanistic studies of ionizing radiation and oxidative mutagenesis: Genetic effects of a single 8-hydroxyguanine (7-hydro-8-oxoguanine) residue inserted at a unique site in a viral genome. *Biochemistry* **29**, 7024–7032 (1990).
15. F. Mazzei, A. Viel, M. Bignami, Role of MUTYH in human cancer. *Mutat. Res.* **743-744**, 33–43 (2013).
16. N. Al-Tassan, N. H. Chmiel, J. Maynard, N. Fleming, A. L. Livingston, G. T. Williams, A. K. Hodges, D. R. Davies, S. S. David, J. R. Sampson, J. P. Cheadle, Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. *Nat. Genet.* **30**, 227–232 (2002).
17. M. L. Thibodeau, E. Y. Zhao, C. Reisle, C. Ch'ng, H. L. Wong, Y. Shen, M. R. Jones, H. J. Lim, S. Young, C. Cremin, E. Pleasance, W. Zhang, R. Holt, P. Eirew, J. Karasinska, S. E. Kalloger, G. Taylor, E. Majounie, M. Bonakdar, Z. Zong, D. Bleile, R. Chiu, I. Birol, K. Gelmon, C. Lohrisch, K. L. Mungall, A. J. Mungall, R. Moore, Y. P. Ma, A. Fok, S. Yip, A. Karsan, D. Huntsman, D. F. Schaeffer, J. Laskin, M. A. Marra, D. J. Renouf, S. J. M. Jones, K. A. Schrader, Base excision repair deficiency signatures implicate germline and somatic MUTYH aberrations in pancreatic ductal adenocarcinoma and breast cancer oncogenesis. *Cold Spring Harb. Mol. Case Stud.* **5**, (2019).
18. Y. Xie, H. Yang, C. Cunanan, K. Okamoto, D. Shibata, J. Pan, D. E. Barnes, T. Lindahl, M. McIlhatton, R. Fishel, J. H. Miller, Deficiencies in mouse Myh and Ogg1 result in tumor predisposition and G to T mutations in codon 12 of the K-ras oncogene in lung tumors. *Cancer Res.* **64**, 3096–3102 (2004).
19. J. Todoric, L. Antonucci, M. Karin, Targeting inflammation in cancer prevention and therapy. Cancer Prev. Res. **9**, 895–905 (2016).
20. L. B. Alexandrov, J. Kim, N. J. Haradhvala, M. N. Huang, A. W. Tian Ng, Y. Wu, A. Boot, K. R. Covington, D. A. Gordenin, E. N. Bergstrom, S. M. A. Islam, N. Lopez-Bigas, L. J. Klimczak, J. R. McPherson, S. Morganella, R. Sabarinathan, D. A. Wheeler, V. Mustonen; PCAWG Mutational Signatures Working Group, G. Getz, S. G. Rozen, M. R. Stratton, P. C. A. W. G. Consortium, The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
21. G. P. Pfeifer, *Mutagenesis* at methylated CpG sequences. Curr. Top. Microbiol. Immunol. **301**, 259–281 (2006).
22. L. B. Alexandrov, Y. S. Ju, K. Haase, P. Van Loo, I. Martincorena, S. Nik-Zainal, Y. Totoki, A. Fujimoto, H. Nakagawa, T. Shibata, P. J. Campbell, P. Vineis, D. H. Phillips, M. R. Stratton, Mutational signatures associated with tobacco smoking in human cancer. *Science* **354**, 618–622 (2016).
23. M. F. Denissenko, A. Pao, M. Tang, G. P. Pfeifer, Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53. *Science* **274**, 430–432 (1996).
24. G. P. Pfeifer, M. F. Denissenko, M. Olivier, N. Tretyakova, S. S. Hecht, P. Hainaut, Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* **21**, 7435–7451 (2002).
25. J. E. Kucab, X. Zou, S. Morganella, M. Joel, A. S. Nanda, E. Nagy, C. Gomez, A. Degasperi, R. Harris, S. P. Jackson, V. M. Arlt, D. H. Phillips, S. Nik-Zainal, A compendium of mutational signatures of environmental agents. *Cell* **177**, 821–836 e816 (2019).
26. S. Nik-Zainal, J. E. Kucab, S. Morganella, D. Glodzik, L. B. Alexandrov, V. M. Arlt, A. Weninger, M. Hollstein, M. R. Stratton, D. H. Phillips, The genome as a record of environmental exposure. *Mutagenesis* **30**, 763–770 (2015).
27. A. Viel, A. Bruselles, E. Meccia, M. Fornasarig, M. Quaia, V. Canzonieri, E. Policicchio, E. D. Urso, M. Agostini, M. Genuardi, E. Lucci-Cordisco, T. Venesio, A. Martayan, M. G. Diodoro, L. Sanchez-Mete, V. Stigliano, F. Mazzei, F. Grasso, A. Giuliani, M. Baiocchi, R. Maestro, G. Giannini, M. Tartaglia, L. B. Alexandrov, M. Bignami, A specific mutational signature associated with DNA 8-oxoguanine persistence in MUTYH-defective colorectal cancer. *EBioMedicine* **20**, 39–49 (2017).
28. C. Mingard, J. Wu, M. McKeague, S. J. Sturla, Next-generation DNA damage sequencing. *Chem. Soc. Rev.* **49**, 7354–7377 (2020).
29. M. Yoshihara, L. Jiang, S. Akatsuka, M. Suyama, S. Toyokuni, Genome-wide profiling of 8-oxoguanine reveals its association with spatial positioning in nucleus. *DNA Res.* **21**, 603–612 (2014).
30. S. Amente, G. Di Palo, G. Scala, T. Castrignano, F. Gorini, S. Cocozza, A. Moresano, P. Pucci, B. Ma, I. Stepanov, L. Lania, P. G. Pelicci, G. I. Dellino, B. Majello, Genome-wide mapping of 8-oxo-7,8-dihydro-2′-deoxyguanosine reveals accumulation of oxidatively-generated damage at DNA replication origins within transcribed long genes of mammalian cells. *Nucleic Acids Res.* **47**, 221–236 (2019).

31. Y. Ding, A. M. Fleming, C. J. Burrows, Sequencing the mouse genome for the oxidatively modified base 8-oxo-7,8-dihydroguanine by OG-seq. *J. Am. Chem. Soc.* **139**, 2569–2572 (2017).
32. A. R. Poetsch, S. J. Boulton, N. M. Luscombe, Genomic landscape of oxidative DNA damage and repair reveals regioselective protection from mutagenesis. *Genome Biol.* **19**, 215 (2018).
33. J. Wu, M. McKeague, S. J. Sturla, Nucleotide-resolution genome-wide mapping of oxidative DNA damage by click-code-seq. *J. Am. Chem. Soc.* **140**, 9783–9787 (2018).
34. B. Cao, X. Wu, J. Zhou, H. Wu, L. Liu, Q. Zhang, M. S. DeMott, C. Gu, L. Wang, D. You, P. C. Dedon, Nick-seq for single-nucleotide resolution genomic maps of DNA modifications and damage. *Nucleic Acids Res.* **48**, 6715–6725 (2020).
35. J. An, M. Yin, J. Yin, S. Wu, C. P. Selby, Y. Yang, A. Sancar, G. L. Xu, M. Qian, J. Hu, Genome-wide analysis of 8-oxo-**7**,8-dihydro-2′-deoxyguanosine at single-nucleotide resolution unveils reduced occurrence of oxidative damage at G-quadruplex sites. *Nucleic Acids Res.* **49**, 12252–12267 (2021).
36. S. G. Jin, D. Pettinga, J. Johnson, P. Li, G. P. Pfeifer, The major mechanism of melanoma mutations is based on deamination of cytosine in pyrimidine dimers as determined by circle-damage-sequencing. *Sci. Adv.* **7**, eabi6508 (2021).
37. S. Boiteux, T. R. O'Connor, F. Lederer, A. Gouyette, J. Laval, Homogeneous Escherichia coli FPG protein. A DNA glycosylase which excises imidazole ring-opened purines and nicks DNA at apurinic/apyrimidinic sites. *J. Biol. Chem.* **265**, 3916–3922 (1990).
38. C. J. Chetsanga, T. Lindahl, Release of 7-methylguanine residues whose imidazole rings have been opened from damaged DNA by a DNA glycosylase from Escherichia coli. *Nucleic Acids Res.* **6**, 3673–3684 (1979).
39. J. Tchou, H. Kasai, S. Shibutani, M. H. Chung, J. Laval, A. P. Grollman, S. Nishimura, 8-oxoguanine (8-hydroxyguanine) DNA glycosylase and its substrate specificity. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4690–4694 (1991).
40. M. A. Kalam, K. Haraguchi, S. Chandani, E. L. Loechler, M. Moriya, M. M. Greenberg, A. K. Basu, Genetic effects of oxidative DNA damages: Comparative mutagenesis of the imidazole ring-opened formamidopyrimidines (Fapy lesions) and 8-oxo-purines in simian kidney cells. *Nucleic Acids Res.* **34**, 2305–2315 (2006).
41. H. J. Helbock, K. B. Beckman, M. K. Shigenaga, P. B. Walter, A. A. Woodall, H. C. Yeo, B. N. Ames, DNA oxidation matters: The HPLC-electrochemical detection assay of 8-oxo-deoxyguanosine and 8-oxo-guanine. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 288–293 (1998).
42. B. Tian, J. Hu, H. Zhang, C. S. Lutz, A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* **33**, 201–212 (2005).
43. M. Jung, G. P. Pfeifer, CpG islands, *in Brenner's Encyclopedia of Genetics (Second Edition)* (Elsevier, 2013), pp. 205–207.
44. V. E. Hoskins, K. Smith, K. L. Reddy, The shifting shape of genomes: Dynamics of heterochromatin interactions at the nuclear lamina. Curr. Opin. Genet. Dev. **67**, 163–173 (2021).
45. L. Guelen, L. Pagie, E. Brasset, W. Meuleman, M. B. Faza, W. Talhout, B. H. Eussen, A. de Klein, L. Wessels, W. de Laat, B. van Steensel, Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
46. X. Zou, G. C. C. Koh, A. S. Nanda, A. Degasperi, K. Urgo, T. I. Roumeliotis, C. A. Agu, C. Badja, S. Momen, J. Young, T. D. Amarante, L. Side, G. Brice, V. Perez-Alonso, D. Rueda, C. Gomez, W. Bushell, R. Harris, J. S. Choudhary; Genomics England Research Consortium, J. Jiricny, W. C. Skarne, S. Nik-Zainal, A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. *Nat. Cancer* **2**, 643–657 (2021).
47. C. Pilati, J. Shinde, L. B. Alexandrov, G. Assie, T. Andre, Z. Helias-Rodzewicz, R. Ducoudray, D. Le Corre, J. Zucman-Rossi, J. F. Emile, J. Bertherat, E. Letouze, P. Laurent-Puig, Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. *J. Pathol.* **242**, 10–15 (2017).
48. M. Ohno, K. Sakumi, R. Fukumura, M. Furuichi, Y. Iwasaki, M. Hokama, T. Ikemura, T. Tsuzuki, Y. Gondo, Y. Nakabeppu, 8-Oxoguanine causes spontaneous de novo germline mutations in mice. *Sci. Rep.* **4**, 4689 (2014).
49. L. Eide, M. Bjoras, M. Pirovano, I. Alseth, K. G. Berdal, E. Seeberg, Base excision of oxidative purine and pyrimidine DNA damage in Saccharomyces cerevisiae by a DNA glycosylase with sequence similarity to endonuclease III from Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10735–10740 (1996).
50. T. K. Hazra, T. Izumi, I. Boldogh, B. Imhoff, Y. W. Kow, P. Jaruga, M. Dizdaroglu, S. Mitra, Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3523–3528 (2002).
51. H. Dou, S. Mitra, T. K. Hazra, Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2. *J. Biol. Chem.* **278**, 49679–49684 (2003).
52. F. Gorini, G. Scala, G. Di Palo, G. I. Dellino, S. Cocozza, P. G. Pelicci, L. Lania, B. Majello, S. Amente, The genomic landscape of 8-oxodG reveals enrichment at specific inherently fragile promoters. *Nucleic Acids Res.* **48**, 4309–4324 (2020).
53. A. Klungland, I. Rosewell, S. Hollenbach, E. Larsen, G. Daly, B. Epe, E. Seeberg, T. Lindahl, D. E. Barnes, Accumulation of premutagenic DNA lesions in mice defective in removal of oxidative base damage. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13300–13305 (1999).
54. H. Sampath, Oxidative DNA damage in disease—Insights gained from base excision repair glycosylase-deficient mouse models. *Environ. Mol. Mutagen.* **55**, 689–703 (2014).
55. J. Todoric, M. Karin, The fire within: Cell-autonomous mechanisms in inflammation-driven cancer. *Cancer Cell* **35**, 714–720 (2019).
56. M. Dizdaroglu, P. Jaruga, M. Birincioglu, H. Rodriguez, Free radical-induced damage to DNA: Mechanisms and measurement. *Free Radic. Biol. Med.* **32**, 1102–1115 (2002).
57. J. L. Ravanat, T. Douki, P. Duez, E. Gremaud, K. Herbert, T. Hofer, L. Lasserre, C. Saint-Pierre, A. Favier, J. Cadet, Cellular background level of 8-oxo-**7**,8-dihydro-2′-deoxyguanosine: An isotope based method to evaluate artefactual oxidation of DNA during its extraction and subsequent work-up. *Carcinogenesis* **23**, 1911–1918 (2002).
58. G. P. Pfeifer, How tobacco smoke changes the (epi)genome. *Science* **354**, 549–550 (2016).
59. L. Lan, S. Nakajima, Y. Oohata, M. Takao, S. Okano, M. Masutani, S. H. Wilson, A. Yasui, In situ analysis of repair processes for oxidative DNA damage in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 13738–13743 (2004).
60. E. Parlanti, M. D'Errico, P. Degan, A. Calcagnile, A. Zijno, I. van der Pluijm, G. T. van der Horst, D. S. Biard, E. Dogliotti, The cross talk between pathways in the repair of 8-oxo-**7**,8-dihydroguanine in mouse and human cells. *Free Radic. Biol. Med.* **53**, 2171–2177 (2012).
61. S. Olafsson, R. E. McIntyre, T. Coorens, T. Butler, H. Jung, P. S. Robinson, H. Lee-Six, M. A. Sanders, K. Arestang, C. Dawson, M. Tripathi, K. Strongili, Y. Hooks, M. R. Stratton, M. Parkes, I. Martincorena, T. Raine, P. J. Campbell, C. A. Anderson, Somatic evolution in non-neoplastic IBD-affected colon. *Cell* **182**, 672–684.e11 (2020).
62. M. L. van den Boogaard, R. Oka, A. Hakkert, L. Schild, M. E. Ebus, M. R. van Gerven, D. A. Zwijnenburg, P. Molenaar, L. L. Hoyng, M. E. M. Dolman, A. H. W. Essing, B. Koopmans, T. Helleday, J. Drost, R. van Boxtel, R. Versteeg, J. Koster, J. J. Molenaar, Defects in 8-oxo-guanine repair pathway cause high frequency of C > A substitutions in neuroblastoma. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2007898118 (2021).
63. A. R. Quinlan, BEDTools: The Swiss-Army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* **47**, 11.12.1–11.12.34 (2014).
64. O. Wagih, ggseqlogo: A versatile R package for drawing sequence logos. *Bioinformatics* **33**, 3645–3647 (2017).
65. J. Ernst, M. Kellis, Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, 2478–2492 (2017).
66. A. Ibarra, C. Benner, S. Tyagi, J. Cool, M. W. Hetzer, Nucleoporin-mediated regulation of cell identity genes. *Genes Dev.* **30**, 2253–2258 (2016).

Acknowledgments

Funding: This work was supported by NIH grant CA228089 to G.P.P. and by the Van Andel Institute. **Author contributions:** S.-G.J., P.E.S., and G.P.P. conceptualized and designed the project. S.-G.J., Y.M., and J.J. performed experiments. S.-G.J., P.E.S., and G.P.P. performed data analysis. G.P.P. and S.-G.J. wrote the manuscript. All authors commented on the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The code for data analysis in this project is archived at DOI: 10.5281/zenodo.6458592. Sequencing data have been deposited into the GEO database (accession number GSE184820).

Submitted 22 November 2021
Accepted 15 April 2022
Published 3 June 2022
10.1126/sciadv.abn3815

PHYSIOLOGY

The histone demethylase KDM5C controls female bone mass by promoting energy metabolism in osteoclasts

Huadie Liu[1]†, Lukai Zhai[2]†, Ye Liu[1], Di Lu[1], Alexandra Vander Ark[2], Tao Yang[1]*, Connie M. Krawczyk[2]*

Women experience osteoporosis at higher rates than men. Aside from hormones, the mechanisms driving sex-dependent bone mass regulation are not well understood. Here, we demonstrate that the X-linked H3K4me2/3 demethylase KDM5C regulates sex-specific bone mass. Loss of KDM5C in hematopoietic stem cells or bone marrow monocytes increases bone mass in female but not male mice. Mechanistically, loss of KDM5C impairs the bioenergetic metabolism, resulting in impaired osteoclastogenesis. Treatment with the KDM5 inhibitor reduces osteoclastogenesis and energy metabolism of both female mice and human monocytes. Our report details a sex-dependent mechanism for bone homeostasis, connecting epigenetic regulation to osteoclast metabolism and positions KDM5C as a potential target for future treatment of osteoporosis in women.

INTRODUCTION

Bone mass in adults is controlled by the coordination and balance between osteoclast (OC)–mediated bone resorption and osteoblast-mediated bone formation. OCs are multinucleated cells of hematopoietic origin formed through the fusion of mononuclear OC precursors from bone marrow monocytes (BMM). Dysregulation of OC-mediated bone resorption has been associated with bone mass–related diseases, such as osteoporosis. Women have lower average bone mass than men, conferring a two- to fourfold increase in osteoporosis and age-related fractures (*1*). Systemic pathways—in particular, the sex hormones—are prominent in regulating sex-dependent bone mass and have been extensively studied (*2*). However, ex vivo–cultured female OC progenitors are more potent in osteoclastogenesis than male OC progenitors, demonstrating a role for factors intrinsic to OCs regulating sexually dimorphic responses (*3*, *4*). However, the identification of which intrinsic factors that mediate sex-dependent bone homeostasis is still lacking.

Sex chromosomes are the fundamental genetic difference between sexes. Discovering how X- or Y-linked genes contribute to sex-dependent bone mass regulation has the potential to lead to the development of promising therapeutics for osteoporosis in women. An increasing number of studies have revealed that OC differentiation and activity are controlled by epigenetic regulation, largely through controlling the accessibility of transcriptional machinery on key OC genes (*5*, *6*). However, until now, the epigenetic factors that regulate OC differentiation and function have been found to work similarly in both sexes (*5*, *6*). KDM5C (JARID1C/SMCX) is an X-linked lysine H3K4 demethylase that escapes X inactivation, resulting in higher expression in females than males (*7*). Males express a paralog of KDM5C from Chr-Y, KDM5D, which is also a H3K4 demethylase (*8*). Global loss of KDM5C results in spurious transcription of genes normally silenced during development (*9*, *10*). KDM5C has been linked to several sex-dependent phenotypes including X-linked intellectual disability (XLID), autism, X inactivation, adiposity, immune response, and cancer (*11*–*17*). In men and male mice, variants of KDM5C cause XLID with short stature, aggressive behavior, and autism (*18*–*20*). Despite the short stature observed in these individuals, a role for KDM5C in bone homeostasis has not been reported. Here, we report that the loss of KDM5C in BMM increases bone mass in female mice, impairs osteoclastogenesis, and reduces bioenergetic metabolism in OC precursors. Thus, KDM5C represents a cell-intrinsic, sex-dependent epigenetic regulator of osteoclastogenesis and bone mass in females, and its inhibition provides a potential therapeutic strategy for preventing osteoporosis in females.

RESULTS

Loss of KDM5C in hematopoietic cells results in increased trabecular bone mass in female mice

To investigate the function of KDM5C in the hematopoietic-OC lineage, we generated *Vav-iCre*; *Kdm5c*$^{fl/fl}$ mice, which diminishes KDM5C expression in all hematopoietic cells including BMM that give rise to OCs. No severe global developmental defects were observed in either sex of the KDM5C-conditionally deficient mice. However, upon closer examination, we noticed a marked increase in macroscopic trabecular bone volume in female *Vav-iCre*; *Kdm5c*$^{fl/fl}$ mice (f*Kdm5c*$^{\Delta Vav}$). When analyzed for bone density and microstructure by microcomputed tomography (micro-CT), bones from f*Kdm5c*$^{\Delta Vav}$ mice were found to have substantially increased trabecular bone mass at 16 weeks of age (Fig. 1A), as indicated by increased trabecular bone volume/total volume (BV/TV), trabecular number (Tb.N), trabecular thickness (Tb.Th), and decreased trabecular separation (Tb.Sp) (Fig. 1B) compared to female control littermates (fCtrl). In contrast, the male *Vav-iCre*; *Kdm5c*$^{fl/fl}$ mice (m*Kdm5c*$^{\Delta Vav/Y}$) and female KDM5C heterozygous conditional knockout mice (f*Kdm5c*$^{\Delta Vav/X}$) have no obvious trabecular bone architecture difference compared to mCtrl and fCtrl, respectively (Fig. 1, A and B). Parameters of cortical bone did not show significant changes between groups (Fig. 1C). These data

[1]Laboratory of Skeletal Biology, Department of Cell Biology, Van Andel Institute, 333 Bostwick Ave NE, Grand Rapids, MI 49503, USA. [2]Department of Metabolism and Nutritional Programming, Van Andel Research Institute, Grand Rapids, MI 49503, USA.
*Corresponding author. Email: tao.yang@vai.org (T.Y.); connie.krawczyk@vai.org (C.M.K.)
†These authors contributed equally to this work.

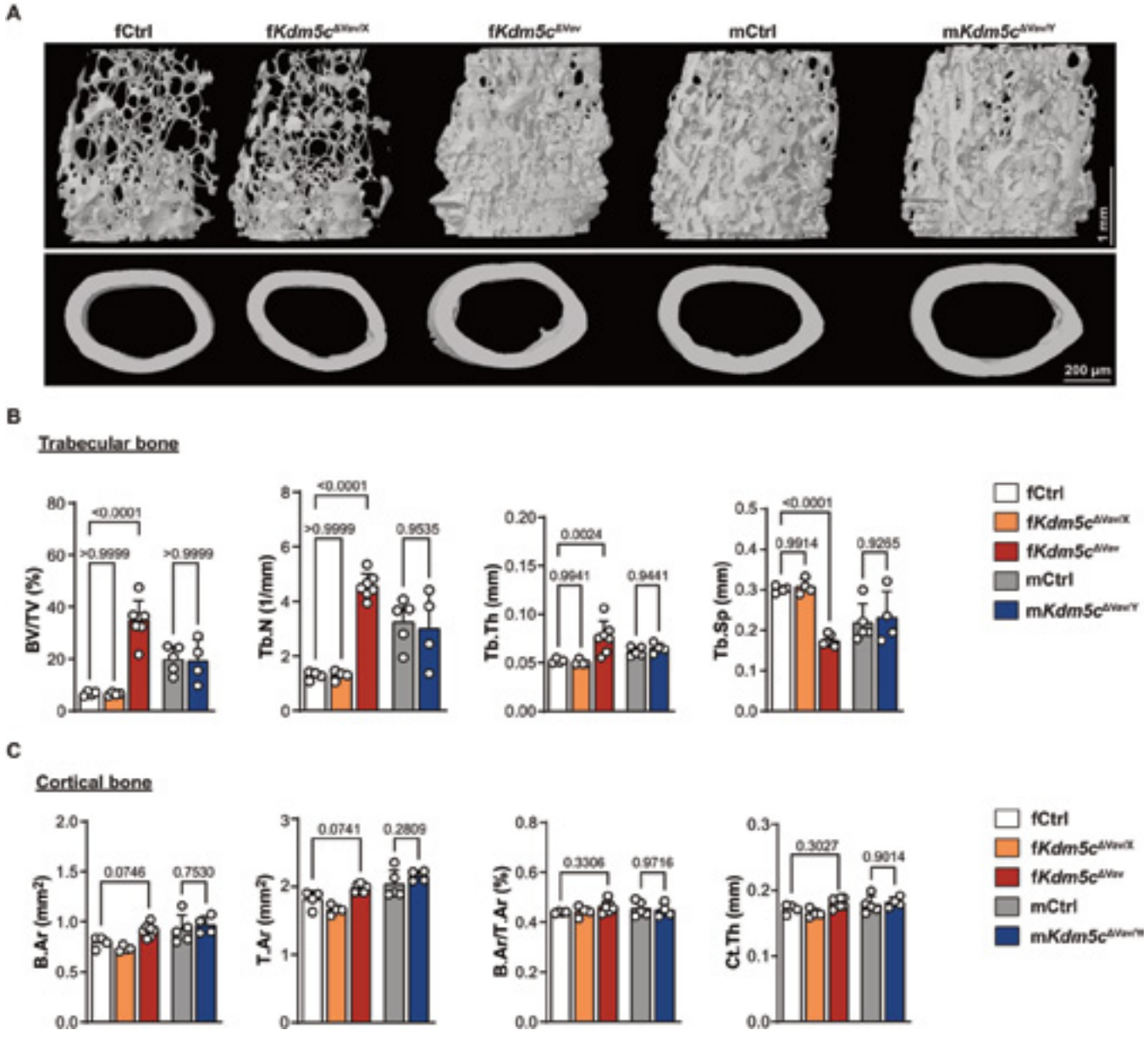


Fig. 1. Increased trabecular bone mass in distal femurs of female *Kdm5c*$^{\Delta Vav}$ but not male *Kdm5c*$^{\Delta Vav}$ mice. (**A**) Representative micro-CT images and (**B** and **C**) quantitation of femur trabecular bone (top; scale bar, 1 mm) and cortical bone (bottom; scale bar, 200 μm) from 16-week-old mice of indicated genotypes/sex. Data comparisons are conducted using one-way analysis of variance (ANOVA). Each data point represents an individual mouse. Data are presented as means ± SEM.

reveal a sex-dependent role for KDM5C in regulating trabecular bone mass.

KDM5C intrinsically regulates osteoclastogenesis

Bone mass in adults is regulated by the coordination and balance between osteoclast-mediated bone resorption and osteoblast-mediated bone formation. We examined bone slices for the presence of OC using tartrate-resistant acid phosphatase (TRAP) staining. We found that the bones from f*Kdm5c*$^{\Delta Vav}$ mice have significantly decreased OC surface/bone surface (Oc.S/BS) and OC number/bone surface (Oc.N/BS) in trabecular bone compared to controls (Fig. 2A), indicating that osteoclastogenesis and bone resorption are impaired. Other hematopoietic cell types such as T and B cells, which participate in bone mass regulation indirectly by affecting the bone microenvironment (*21*), would also be affected in the *Vav-iCre* mice. To elucidate whether KDM5C regulates bone formation intrinsically in the myeloid lineage, including monocytes and OCs, we isolated BMM from bone marrow of control and f*Kdm5c*$^{\Delta Vav}$ mice and analyzed their ability to generate OCs ex vivo following RANKL (receptor activator of nuclear factor κB ligand) treatment. OC formation was significantly impaired in f*Kdm5c*$^{\Delta Vav}$ BMM, as indicated by the reduced Oc.N and Oc.S (Fig. 2B) and decreased expression of osteoclastogenic gene transcripts (Fig. 2C). Furthermore, we generated *LysM-Cre*; *Kdm5c*$^{fl/fl}$

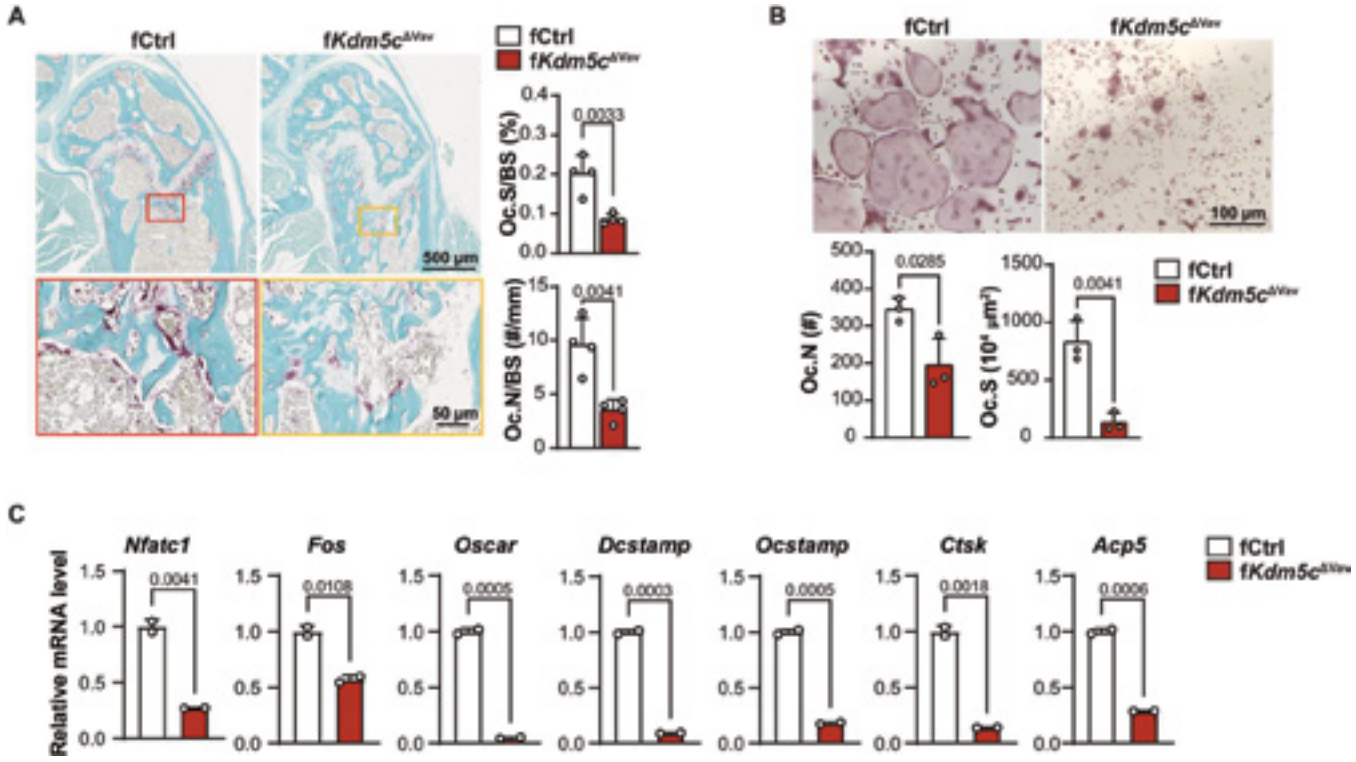


Fig. 2. Impaired osteoclastogenesis in female *Kdm5c*$^{\Delta Vav}$ mice. (**A**) TRAP staining of sectioned femurs from 16-week-old fCtrl and f*Kdm5c*$^{\Delta Vav}$ mice. Oc.N/BS and Oc.S/BS in trabecular bone area were measured and calculated (n = 4 per genotype). Scale bars, 500 μm (top) and 50 μm (bottom). (**B**) Ex vivo osteoclastogenesis on BMM of fCtrl and f*Kdm5c*$^{\Delta Vav}$ mice. OCs were visualized by TRAP staining 3 to 5 days after differentiation. Oc.N and Oc.S were measured and calculated (n = 3 per genotype). Scale bar, 100 μm. (**C**) mRNA levels of OC/osteoclastogenic genes in fCtrl and f*Kdm5c*$^{\Delta Vav}$ BMM 48 hours after osteoclastogenic induction. mRNA levels were detected by quantitative reverse transcription polymerase chain reaction (qRT-PCR). Relative mRNA levels are shown (n = 2 per genotype). All data comparisons are conducted by Student's t test, two-tailed. Data are presented as means ± SEM.

mice, which have *Kdm5c* deleted in myeloid cells. Similar to f*Kdm5c*$^{\Delta Vav}$ mice, female *LysM-Cre*; *Kdm5c*$^{fl/fl}$ mice (f*Kdm5c*$^{\Delta LysM}$) had significantly increased trabecular bone mass and decreased OC activity, as indicated by increased BV/TV and Tb.N and decreased Tb.Sp, Oc.S/BS, and Oc.N/BS (Fig. 3, A and B). Consistent with the f*Kdm5c*$^{\Delta Vav}$ mice, we observed no significant differences in cortical bone mass (Fig. 3A). Overall, the bone phenotypes in the f*Kdm5c*$^{\Delta LysM}$ were less severe compared to the f*Kdm5c*$^{\Delta Vav}$ mice. We found that BMM isolated from f*Kdm5c*$^{\Delta LysM}$ mice show significant reduction in osteoclastogenic potential ex vivo, however not to the same extent as observed in the *Kdm5c*$^{\Delta Vav}$ BMM (Fig. 3, C and D). This is likely due to reduced KDM5C deletion efficiency in f*Kdm5c*$^{\Delta LysM}$ BMM compared to f*Kdm5c*$^{\Delta Vav}$ BMM (fig. S1, A to D). Together, these findings demonstrate that loss of KDM5C intrinsically in myeloid progenitor cells results in reduced osteoclastogenesis.

Next, to investigate whether the intrinsic regulation of KDM5C in osteoclastogenesis is sex dependent, we compared the osteoclastogenic ability between mCtrl (m*Kdm5c*$^{wt/Y}$ *Cre*$^{+}$) and m*Kdm5c*$^{\Delta Vav/Y}$ BMM. KDM5C-deficient BMM had decreased osteoclastogenesis compared to controls (fig. S2A). The expression of OC genes (*Nfatc1*, *Dcstamp*, and *Acp5*) also decreased in m*Kdm5c*$^{\Delta Vav/Y}$ cells compared with mCtrl, however to a lesser degree compared to female cells (fig. S2B). This suggests that a compensatory mechanism (e.g., KDM5D) is present in male BMM for osteoclastogenesis. Thus, KDM5C controls bone mass in a sex-specific manner due to its differential ability to control osteoclastogenesis in male versus female BMM.

Mitochondrial respiration and energy metabolism are impaired following the loss of KDM5C in BMM

KDM5C not only represses transcription via demethylation of H3K4me2/3 at gene promoters but also promotes gene expression in mouse embryonic stem cells (*10*, *15*) and ERα-positive breast cancer cells (*22*) by converting H3K4me2/3 modifications into H3K4me1 or recruiting transcription factors on specific transcriptional enhancers. Therefore, to investigate how KDM5C affects transcriptional programming of BMM, we compared gene expression of fCtrl and f*Kdm5c*$^{\Delta Vav}$ BMM at different stages of osteoclastogenesis ex vivo (0, 16, and 32 hours). Significant transcriptomic differences were observed between fCtrl and f*Kdm5c*$^{\Delta Vav}$ at all three stages (fig. S3). Using Gene Ontology analysis, we found that genes that were increased in KDM5C-deficient BMM/OCs were related to immune cell activity and inflammation, while genes with decreased expression were enriched in metabolic and mitochondrial respiration–related pathways (Fig. 4A). Notably, mitochondrial adenosine triphosphate (ATP) synthesis and electron transport pathways are repressed the most in KDM5C-deficient cells at early stages of OC formation (Fig. 4B, 32 hours). These data demonstrate that

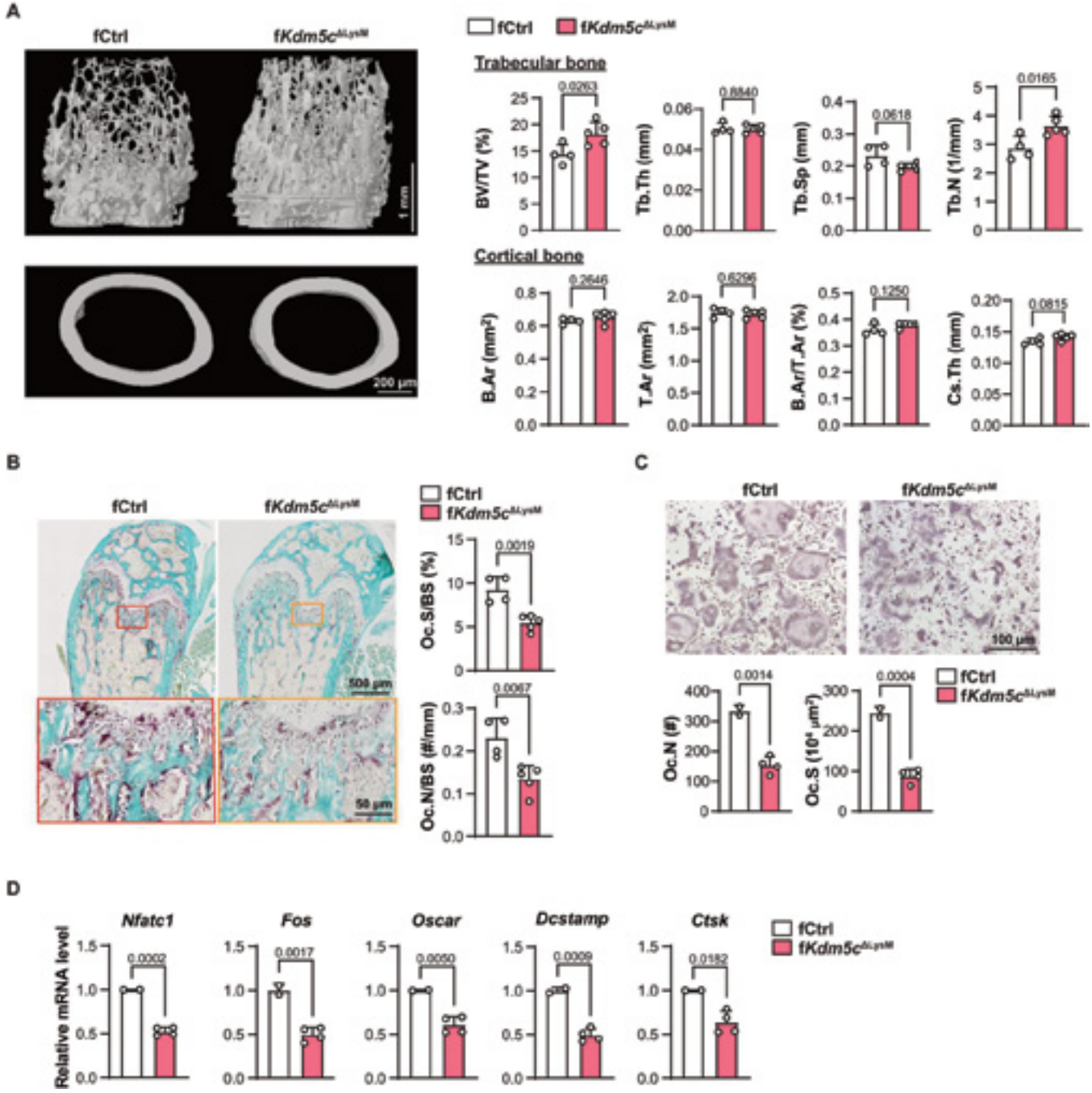


Fig. 3. Increased trabecular bone mass and impaired osteoclastogenesis in f*Kdm5c*$^{\Delta LysM}$ mice. (**A**) Representative micro-CT images and quantitation of femur trabecular bone (top; scale bar, 1 mm) and cortical bone (bottom; scale bar, 200 μm) from 8-week-old fCtrl and f*Kdm5c*$^{\Delta LysM}$ mice. (**B**) Representative TRAP staining images of sectioned femurs from 8-week-old fCtrl and f*Kdm5c*$^{\Delta LysM}$ mice, used to calculate Oc.N/BS and Oc.S/BS ($n = 4$ for fCtrl; $n = 5$ for f*Kdm5c*$^{\Delta LysM}$). Scale bars, 500 μm (top) and 50 μm (bottom). (**C**) Ex vivo osteoclastogenesis of BMM from fCtrl and f*Kdm5c*$^{\Delta LysM}$ mice. OCs were visualized by TRAP staining 3 to 5 days after differentiation. Oc.N and Oc.S were measured and calculated ($n = 3$ per genotype). Scale bar, 100 μm. (**D**) mRNA levels of OC/osteoclastogenic genes were detected by qRT-PCR in fCtrl and f*Kdm5c*$^{\Delta LysM}$ BMM 48 hours after induction. Relative mRNA levels are shown ($n = 2$ for fCtrl, $n = 4$ for f*Kdm5c*$^{\Delta LysM}$). All data comparisons are conducted by Student's *t* test, two-tailed. Data are presented as means ± SEM.

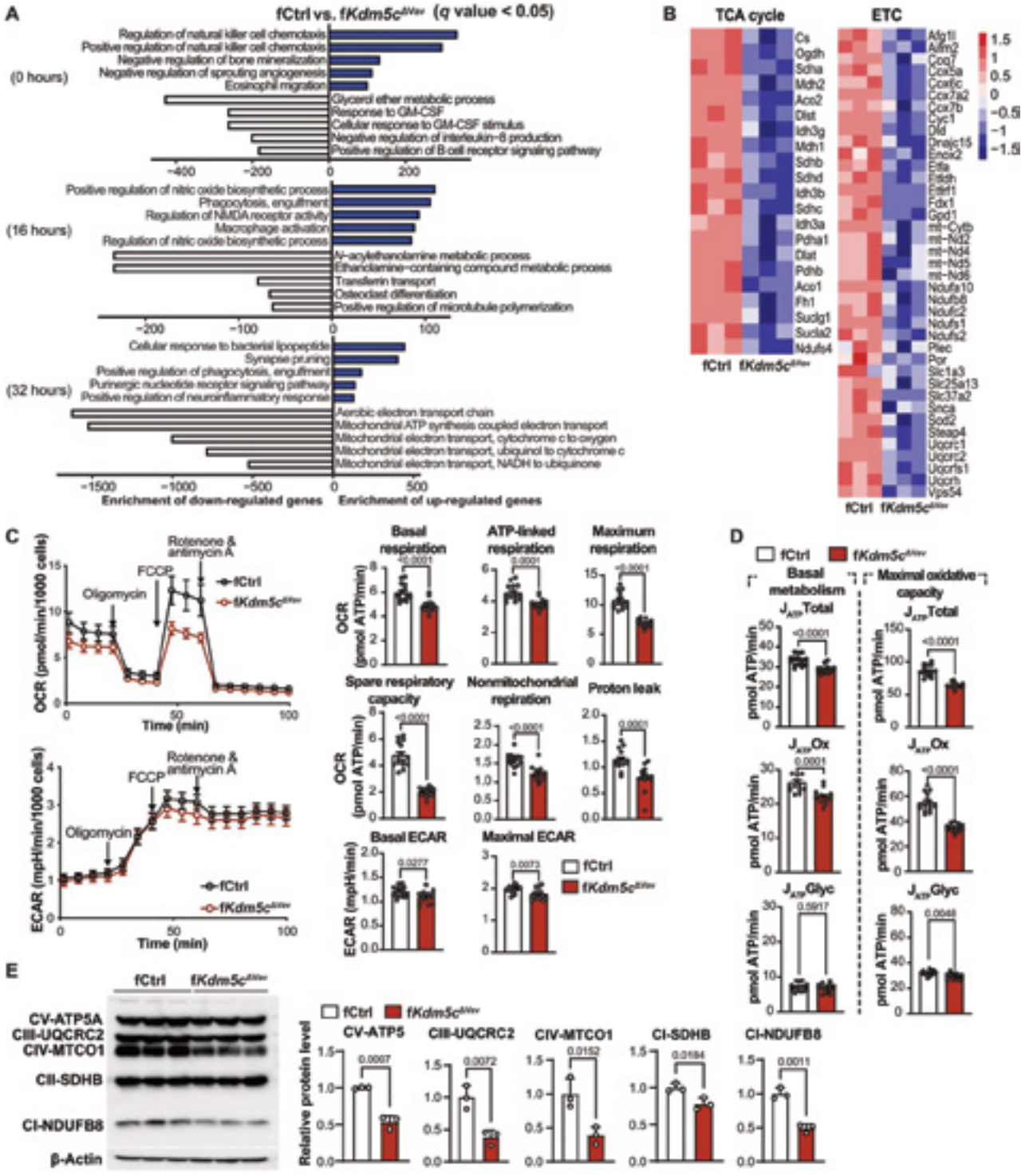


Fig. 4. KDM5C-deficient BMM have decreased bioenergetic metabolism during osteoclastogenesis. (**A**) The top five enriched gene ontology biological process terms assigned to up-regulated (blue bars) or down-regulated (white bars) genes in f*Kdm5c*$^{\Delta Vav}$ versus fCtrl cells during osteoclastogenesis. Gene expression was determined by RNA-seq at three time points during osteoclastogenesis. Genes with false discovery rate < 0.05 were chosen for analysis. (**B**) Heatmap of TCA cycle genes and mitochondrial electron transport chain (ETC) genes expression in f*Kdm5c*$^{\Delta Vav}$ and control BMM after osteoclastogenic induction. (**C**) OCR (top) and ECAR (bottom) in fCtrl and f*Kdm5c*$^{\Delta Vav}$ BMM. OCR and ECAR were detected 72 hours after osteoclastogenic induction. Detailed parameters including basal and maximum respiration, ATP-coupled respiration, spare respiratory capacity, nonmitochondrial respiration, and proton leak are shown in the bottom. (**D**) Oxidative ATP production rate and maximum glycolytic ATP production in f*Kdm5c*$^{\Delta Vav}$ and control BMM. (**E**) Immunoblot of mitochondrial OXPHOS complex proteins (left) and densitometry quantifications (right) of control and *Kdm5c*-deficient BMM stimulated by RANKL for 72 hours. Each lane represents one mouse ($n = 3$ per genotype). All data comparisons are conducted by Student's *t* test, two-tailed. Data are presented as means ± SEM.

KDM5C promotes the expression of genes related to mitochondrial metabolism in BMM/OC.

Osteoclastogenesis and osteolysis are energy-consuming processes (*23*), and preosteoclast differentiation and OC survival are suppressed when mitochondrial metabolism is impaired (*23*–*25*). We found that mitochondrial metabolism genes increased significantly over the course of OC differentiation, supporting the essential role of mitochondrial metabolism in osteoclastogenesis (fig. S4, A and B). Therefore, to test whether decreased gene expression alters mitochondrial bioenergetic metabolism, we measured oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) as surrogates of mitochondrial respiration and glucose fermentation (conversion of pyruvate to lactate), respectively (Fig. 4C). Notably, basal respiration, ATP-coupled respiration, spare respiratory capacity, and maximal respiratory capacity were all significantly reduced in KDM5C-deficient BMM compared to controls (Fig. 4C). We also found that KDM5C-deficient BMM have reduced glucose fermentation (Fig. 4C). These results demonstrate that KDM5C positively regulates bioenergetic metabolism. This was confirmed by the decrease in ATP production from oxidative phosphorylation (OXPHOS) and glycolysis (Fig. 4D). Our results also confirm previous results that OCs generate most of their ATP via OXPHOS (Fig. 4D) (*24*). We examined the expression of proteins involved in electron transport chain (ETC) assembly and function and found that ATP5, UQCRC2, MTCO1, SDHB, and NDUFB8 were all reduced in the KDM5C-deficient cells relative to controls (Fig. 4E and fig. S5A). Consist with *Vav-iCre* genes involved in ETC assembly and function including *Atp6vod2*, *Uqcrc2*, *Atp1b3*, *Cox6b1*, *Sdhb*, and *Sdha* had lower expression in *LysM-Cre* KDM5C–deficient BMM (fig. S5B). Similar to OC genes (fig. S2B), the expression of mitochondrial genes was decreased in male KDM5C–deficient cells but to a lesser extent than observed in females (fig.S5D). To examine whether specific complexes of the ETC were differentially essential for osteoclastogenesis, we inhibited complex I, II, or III, using piericidin, atpenin A5, or myxothiazol/antimycin, respectively. We found that inhibition of any of the complexes completely blocked OC formation with no TRAP-positive cells compared to control group [dimethyl sulfoxide (DMSO) treatment] (fig. S6). These results indicate that mitochondrial ETC is essential for osteoclastogenesis. Overall, these data demonstrate that KDM5C is necessary for bioenergetic metabolism in female BMM/OC, largely through promoting mitochondrial respiration.

KDM5C-regulated osteoclastogenesis is mediated in part by PGC-1β

Next, we investigated how KDM5C loss alters metabolic programming in the BMM/OC population. PGC-1α and PGC-1β are transcriptional coactivators that promote mitochondrial biogenesis and the expression of mitochondrial metabolism genes (*26*). Previous studies reported that PGC-1β supports OC formation (*27*, *28*) by enhancing mitochondrial biogenesis and cytoskeletal rearrangement (*29*). We found that PGC-1β (*Ppargc1b*), but not PGC1-α (*Ppargc1a*), was highly expressed in control BMM and increased during osteoclastogenesis (Fig. 5A). We found that PGC-1β expression did not increase to the same extent in KDM5C-deficient BMM/OC (Fig. 5A and fig. S5C). To determine whether increasing the expression of PGC-1β in KDM5C-deficient BMM could rescue osteoclastogenesis, we overexpressed PGC-1β in control and KDM5C-deficient BMM using retroviral expression. We found that the osteoclastogenic potential of KDM5C-deficient BMM could be partially restored by PGC-1β, as indicated by increased Oc.N and Oc.S in an ex vivo osteoclastogenesis assay (Fig. 5B). PGC-1β–mediated regulation of mitochondrial metabolism has been shown to be dependent on iron uptake by the transferrin receptor protein 1 (TfR1) (*27*), and iron itself is required for synthesis of cofactors essential to the function of enzymes in the ETC. We examined the expression of *Tfrc* (the gene encoding TfR1) in KDM5C-deficient BMM/OC and found decreased expression relative to controls (fig. S7). These data show that PGC-1β mediates KDM5C-regulated osteoclastogenesis, but other factors, including iron uptake via TfR1 and/or the expression of other mitochondrial genes, are also likely contributing.

KDM5 inhibition dampens osteoclastogenic potential of both mouse and human monocytes

Next, we tested whether we could prevent osteoclastogenesis in fCtrl BMM by using small-molecule inhibitors of KDM5C demethylase activity. While no KDM5C-specific inhibitor exists, pan-KDM5 inhibitors (KDM5i) are available and have been used in clinical trials for the treatment of cancer and hepatitis B (*30*, *31*). Our RNA sequencing (RNA-seq) data show that *Kdm5c* is expressed more highly than other KDM5 family members in female BMM (Fig. 6A), suggesting that KDM5C may have a greater role than other KDM5 members in regulating osteoclastogenesis. We found that KDM5i dose-dependently suppressed RANKL-induced osteoclastogenesis of female BMM, indicated by TRAP staining in an ex vivo osteoclastogenesis assay, with a median inhibitory concentration (IC_{50}) of 5.6 μM (Fig. 6B). Consistently, KDM5i treatment significantly down-regulated key mitochondria OXPHOS complex proteins (Fig. 6C) and mRNAs (*Sdhb*, *Atp6v0d2*, *Sdha, Uqcrc2*, and *Uqcrfs1*), as well as mRNA expression of OC marker genes (*Dcstamp*, *Acp5*, and *Ctsk*) and *Ppargc1b* (fig. S8A). OCR, ECAR, and ATP production were also impaired by KDM5i in a dose-dependent manner (Fig. 6, D and E), demonstrating that the function of KDM5C in regulating mitochondrial metabolism is dependent on its demethylase activity. KDM5i increased H3K4me3 level in $CD14^+$ monocytes of human peripheral blood mononuclear cells (PBMCs) (fig. S8B). The effect of KDM5i on osteoclastogenesis and energy metabolism is conserved in female mice and humans, indicated by the reduction in the number and area of multinuclear OCs (Fig. 6F) and the impairment in OCR, ECAR, and ATP production (Fig. 6G and fig. S8C) in KDM5i-treated female peripheral human blood monocytes under osteoclastogenic differentiation. Thus, targeting KDM5C has the potential to reduce OC function therapeutically, potentially mitigating bone loss in females.

DISCUSSION

Our work has identified KDM5C, an X-linked chromatin-modifying enzyme, as a female-specific regulator of bone mass that promotes OC differentiation and function. Females, without KDM5C in BMM/OC, displayed increased bone mass comparable to control or KDM5C-deficient males demonstrating that KDM5C uniquely regulates bone mass in females. Mechanistically, we found that KDM5C promotes bioenergetic metabolism required for osteoclastogenesis. Thus, we have identified a mechanism linking epigenetic and metabolic programming of OCs to sex-specific bone mass regulation. Females are at a higher risk than males for developing osteoporosis, mainly because of the inherent lower peak bone mass

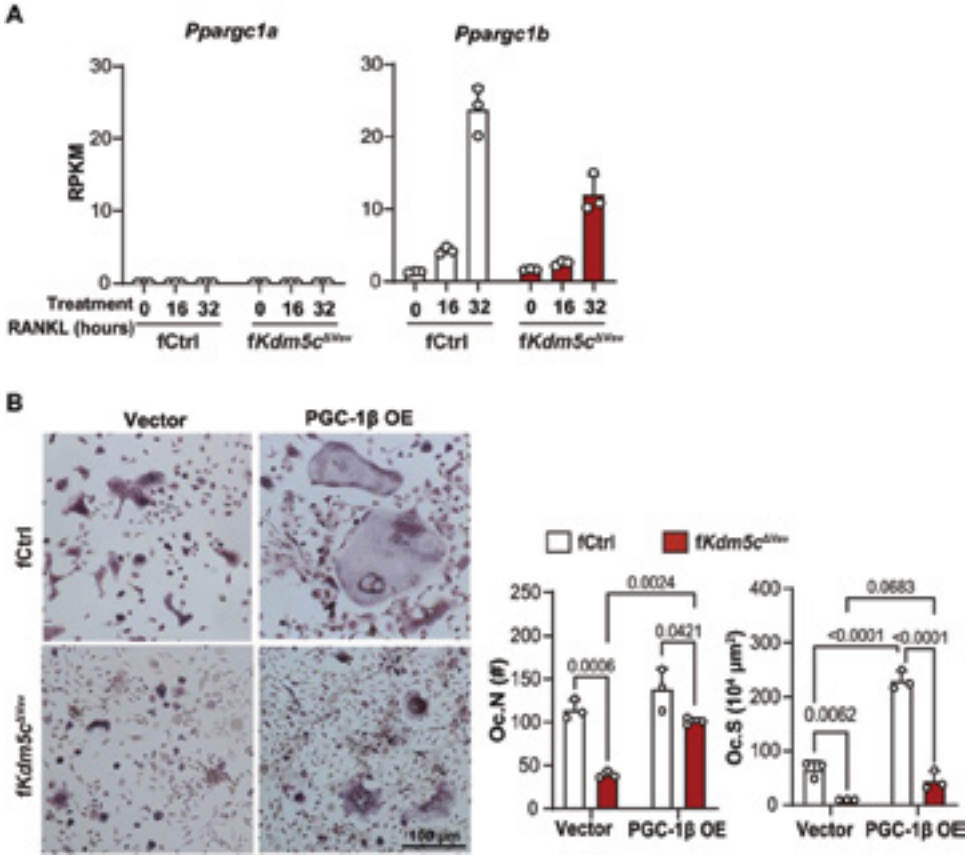


Fig. 5. Partial rescue of osteoclastogenesis by PGC-1β. (**A**) *Ppargc1b* and *Ppargc1a* expression (RNA-seq) during ex vivo osteoclastogenesis assay at indicated time points. (**B**) Ex vivo osteoclastogenesis measured by TRAP staining of BMM from fCtrl and f*Kdm5c*$^{\Delta Vav}$ mice transduced with control or PGC-1β overexpression vector (two-way ANOVA analyses). Scale bar, 100 μm. Data are presented as means ± SEM.

before menopause and a more aggressive bone loss at postmenopause stage. Estrogen not only inhibits OC formation and function and was used for osteoporosis treatment for women but also leads to strong side effects, including cancers. Current treatment alternatives to estrogen have been developed and successfully used to treat osteoporosis; however, none of these therapeutics target a mechanism specific to females. Our study showed that pharmacological inhibition of KDM5C blocks osteoclastogenesis of both human and mouse monocytes, indicating that KDM5C is a viable therapeutic target for osteoporosis, particularly for females.

While KDM5C predominantly functions as a transcriptional repressor by removing active H3K4me2/3 marks from promoters, it can also stimulate gene expression dependent on or independent of demethylase activity (*10*, *15*, *22*). Our data show that KDM5C promotes mitochondrial metabolism and related gene expression. In *Drosophila*, the KDM5A-C ortholog KDM5 (Lid) also promotes the transcription of genes important for mitochondrial function (*32*). KDM5 regulation of mitochondrial gene expression is mediated by the PHD3 domain, a domain that is present in KDM5A/B but not KDM5C/D (*32*). In our study, use of the KDM5-specific histone lysine demethylase inhibitor reduces mitochondrial metabolism and related gene expression, demonstrating that KDM5C demethylase activity is critical for BMM/OC bioenergetic metabolism. Demethylase-dependent positive regulation of gene expression by KDM5C occurs by trimming H3k4me2/3 for optimal enhancer activity (*10*). Thus, KDM5C may control the mitochondrial transcriptional program in OCs through enhancer regulation. In particular, *Ppargc1b*, which encodes PGC-1β, a master regulator of mitochondrial metabolism and biogenesis, is significantly downregulated. While we found that PGC-1β ectopic expression partially rescued osteoclastogenesis, PGC-1β is not a rate-limiting mediator of KDM5C-mitochondrial metabolism on its own. Notably, expression of the iron transporter TfR1 (*Tfrc*) was reduced in the absence of KDM5C; this suggests that the KDM5C-dependent iron uptake mechanism and PGC-1β expression may synergize to promote mitochondrial metabolism and osteoclastogenesis (*27*).

Our data showed that KDM5C-deficient male cells also have reduced osteoclastogenesis and expression of mitochondrial metabolism genes but not to the same extent as females. However, these differences did not translate into a difference in bone mass. These data suggest that there may be a gene dosage effect of KDM5C on gene expression and that KDM5D may play an important role in osteoclastogenesis and bone mass in males. KDM5D is also a lysine histone demethylase; however, it is less well-studied compared to KDM5C.The functional discrepancy between KDM5C and KDM5D paralogs was also found in other settings. For example, males with X-linked immunodeficiency caused by KDM5C variants have a more severe phenotype compared to heterozygote females, although each of them carries one functional copy of KDM5D and KDM5C, respectively. These observations

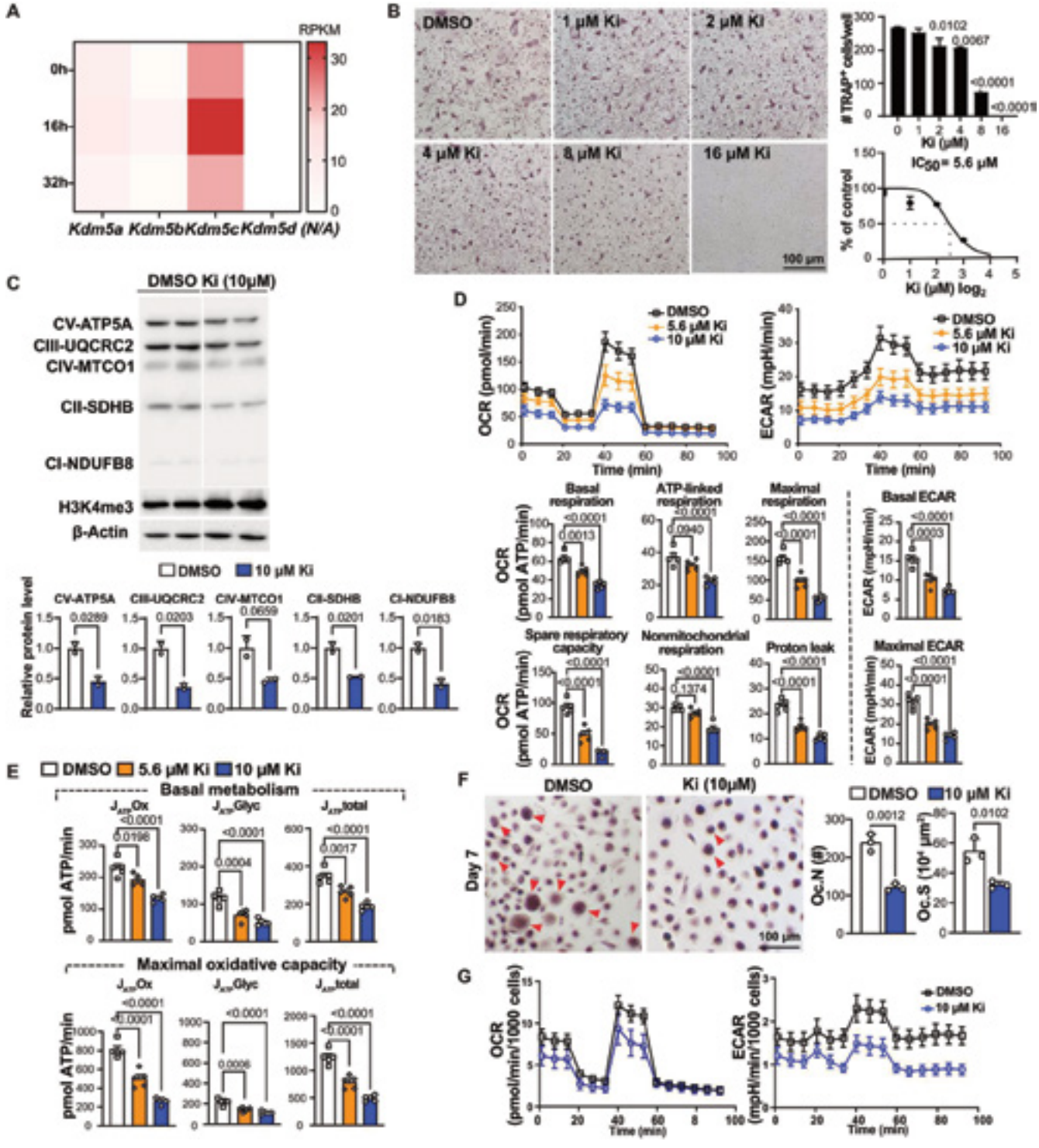


Fig. 6. Pharmacological inhibition of KDM5 impairs mitochondrial metabolism and osteoclastogenesis of mice and human monocytes. (**A**) The gene expression of KDM5 family members during osteoclastogenesis (analyzed by RNA-seq, as in Fig. 5A). (**B**) Representative TRAP staining images of BMM osteoclastogenesis following treatments with different doses of KDM5 inhibitor (Ki) (top). The number of TRAP-positive cells per well was quantified and used to calculate IC_{50} of KDM5i in osteoclastogenesis (bottom). Scale bar, 100 μm. (**C**) Western blots of mitochondrial OXPHOS proteins and H3K4me3 after Ki (10 μM) treatment. (**D**) OCR (left) and ECAR (right) of RANKL-stimulated (72 hours) BMM treated with DMSO, 5.6 or 10 μM Ki. All parameters were calculated as in Fig. 4. (**E**) Oxidative and glycolytic ATP production of BMM following Ki treatment. (**F**) Human PBMC osteoclastogenesis with Ki treatment. Monocytes cultured from human PBMCs were treated with RANKL for 7 days before TRAP staining. Multinuclear (>3) OCs are indicated by red arrowheads. Total numbers (Oc.N) and total area (Oc.S) of OCs were quantified using ImageJ (n = 3 for each group). Scale bar, 100 μm. (**G**) OCR (left) and ECAR (right) of human PBMC monocytes treated with DMSO and 10 μM Ki 5 days after osteoclastogenic induction. Comparisons in (C) and (F) are conducted by Student's t test, two-tailed; in (D) and (E), by one-way ANOVA analyses. Data are presented as means ± SEM.

suggest that KDM5C and KDM5D are not completely redundant (*33*). The unique and redundant roles of KDM5C/D and gene dose effects in bone mass regulation deserve further studies. We also did not see short stature on our KDM5C-deficient male mice, as seen in patients and cytomegalovirus-Cre-drive deletion of KDM5C (*18*), suggesting that the role of KDM5C in height is in cells of nonhematopoietic origin.

Osteoclastogenesis requires robust mitochondrial metabolism (*23*, *24*). We found that expression of mitochondrial genes increases during OC differentiation. PGC1-β promotes mitochondrial biogenesis and, coupled with iron transported by TfR1, increases mitochondrial respiration. PGC1-β and mitochondrial respiration also regulate actin cytoskeletal organization and therefore could also be involved in organization of glycolytic pathway enzymes (*23*, *29*). Further, it has been shown that estrogens decrease Oc.N by reducing OXPHOS in OC precursors (*25*). Glucose metabolism, both fermentation to lactate and oxidation in tricarboxylic acid (TCA) cycle, is also necessary for OC differentiation (*24*). While our study here was focused on ETC activity, it is possible that decreased flux of metabolites through the TCA cycle also contributes to decreased OXPHOS and ATP production. We did observe a decrease in glucose fermentation (measured by ECAR), suggesting that glucose metabolism is reduced. Future work examining the fates of major carbon sources such as glucose, glutamine, and amino acids will determine to what extent the metabolic defects we have observed are due to altered metabolic wiring versus transcriptional programming.

KDM5C gene dose has been associated with increased adiposity in females due to increased expression (*13*). Link *et al.* (*13*) found a strong correlation in humans between body mass index (BMI) and KMD5C expression and associations of noncoding variants with adiposity traits. It is not known whether KDM5C affects cellular metabolism in adipocytes. However, these findings, together with ours, suggest that KDM5C may have a broader role in metabolism beyond osteoclastogenesis.

Here, we find evidence that KDM5C is essential for monocyte differentiation to OCs. Overall, our findings highlight an epigenetic mechanism that controls osteoclastogenesis by governing the transcriptional programming of energy metabolism, positioning KDM5C as a potential target for the treatment of osteoporosis in females.

MATERIALS AND METHODS

Mice

The *Kdm5c*$^{fl/fl}$ mouse with C57BL/6 background was a gift from Y. Shi's laboratory. Briefly, exons 11 and 12 of *Kdm5c* were flanked by *flox* sequences to enable excision by Cre recombinase (*10*, *17*). *Kdm5c*$^{fl/fl}$ mice were crossed to *Vav-iCre* (strain 008610) and *LysM-Cre* (strain 004781) mice, respectively, purchased from The Jackson Laboratory to generate KDM5C knockout mice. *Vav-iCre* is turned on at embryonic day 11.5 in almost all hematopoietic cells (*33*). Whereas *LysM-Cre* is turned on in myeloid progenitors at the granulocyte-monocyte progenitor stage (*34*). No reports have ever shown that express which of these two Cre would affect bone phenotypes (*35*–*37*). All mice were maintained following guidelines of the Institutional Animal Care and Use Committee at Van Andel Institute. Mice euthanization was conducted according to American Veterinary Medical Association Guidelines for the Euthanasia of Animals.

Bone histomorphometry

Bone histomorphometry was performed as described before (*38*). Briefly, paraffin-embedded femurs from 2-month-old mice were sent to Pathology and Biorepository Core at Van Andel Research Institute for sectioning. TRAP staining on paraffin sections were performed using an acid phosphatase, leukocyte kit (Sigma-Aldrich, 387A) by following the standard protocol. N.Oc/BS, Oc.S/BS, and Oc.S/N.Oc were quantified using BioQuant osteo software (Nashville, TN, USA).

Microcomputed tomography

Mouse femurs were scanned following the Journal of Bone and Mineral Research (JBMR)-recommended micro-CT guidelines as previously described (*39*). Reconstruction of cross-sectional slices was performed using NRecon software (SkyScan), while bone parameters were calculated through CTAN and CTVOL software. Bone parameters calculated are BV/TV, Tb.Th, Tb.Sp, Tb.N, cortical bone area (Ct.Ar), cortical thickness, the Ct.Ar/total cross-sectional area (Tt.Ar) ratio, and Tt.Ar.

BMM isolation and culture

Bone marrow cells were collected from femurs and tibias and plated in complete BMM media [α-minimum essential medium plus 10% fetal bovine serum, 1% penicillin-streptomycin, 1% L-glutamine, and macrophage colony-stimulating factor (30 ng/ml)]. The next day, nonadherent cells were transferred to new dishes for BMM culture with fresh complete BMM media and expanded for 48 hours.

Human PBMC extraction and culture

Human blood collection was approved by the Internal Review Board (IRB) of Calvin University (reference number: 20-023). All participating donors have understood and signed IRB-approved consent forms. PBMCs were obtained by centrifugation of blood through a Ficoll-Hypaque density gradient at 300*g* for 60 min. Adherent cells from PBMCs were cultured in BMM media for 1 to 3 days before use.

Ex vivo osteoclastogenesis assay

BMM or cultured human monocytes (3×10^4 or 6×10^4 cells per well, respectively) were seeded into 96-well plates and cultured overnight with complete BMM media. Differentiation was induced by osteoclastogenic medium (50 ng/ml RANKL in complete BMM media), and induction medium was changed every other day. Cells were subjected to staining when OC fusion appeared (3 to 5 days for BMM and 7 to 9 days for human monocytes). TRAP staining was conducted to visualize OCs using an Acid Phosphatase Leukocyte (TRAP) Kit (Sigma-Aldrich) by following the standard protocol. Cells positive for TRAP staining and contain >3 nuclei were counted as OCs. Osteoclast numbers and area were quantified using ImageJ.

Total RNA isolation and qRT-PCR

Total RNA from BMM were extracted using TRIzol reagent (Invitrogen) by following the standard protocol. Genomic DNAs were removed by using the DNA-free DNA Removal kit (Thermo

Fisher Scientific, AM1906) according to the manufacturer's instruction. Five hundred nanograns of RNA was subjected to the synthesis of first-strand cDNA using the SuperScript VILO cDNA synthesis Kit (Invitrogen, 11754050). Quantitative polymerase chain reaction (qPCR) was performed on a StepOne PCR instrument using SYBR Green qPCR Master Mix (Invitrogen, 4472908). Primers used for quantitative reverse transcription PCR (qRT-PCR) in this study are as follows: *Kdm5c*: 5′-GAGCAGTCTGTACTGTGCCA-3′ (forward), 5′-ATTCCCACATACAGCCACGG-3′ (reverse); *Nfatc1*: 5′-TCATCGGCGGGAAGAAGATG-3′ (forward), 5′-GTCCCGGTCAGTCTTTGCTT-3′ (reverse); *Fos*: 5′-TACTACCATTCCCCAGCCGA-3′ (forward), 5′-GCTGTCACCGTGGGGATAAA-3′ (reverse); *Oscar*: 5′-CGTGCTGACTTCACACCAAC-3′ (forward), 5′-GGTCACGTTGATCCCAGGAG-3′ (reverse); *Dcstamp:* GCTGTATCGGCTCATCTCC*T*-3′ (forward), 5′-ATGGACGACTCCTTGGGTTC-3′ (reverse); *Ocstamp:* 5′-AGCCACGGAACACCTCTTTG-3′ (forward), 5′-TGGAACAACTGCCTTGCAGA-3′ (reverse); *Ctsk*: 5′-GAAGGGAAGCAAGCACTGGA-3′ (forward), 5′-CCATGTTGGTAATGCCGCAG-3′ (reverse); *Acp5*: 5′-CTGGTATGTGCTGGCTGGAA-3′ (forward), 5′-CGCAAACGGTAGTAAGGGCT-3′ (reverse); *Atp6v0d2*: 5′-GGGCCTGGTTCGAGGATG-3′ (forward), 5′-GAAGTTGCCATAGTCCGTGGT-3′ (reverse); *Uqcrc2*: 5′-GCAACTGCTAGAGCCATGAAG-3′ (forward), 5′-TTAACCTTCGGGGCAACTTTGA-3′ (reverse); *Apt1b3*: 5′-TGCTGGAACCAGGAACCTAAA-3′ (forward), 5′-CTAGGCTCGTGCTGTGACTT-3′ (reverse); *Cox6b1*: 5′-CGCTACTCCGGGACAATCTT-3′ (forward), 5′-TCTGGTTCTGGTTGGGGAAG-3′ (reverse); *Sdhb*: 5′-CGTTCTCGGCAGAGTCGG-3′ (forward), 5′-GGTCCCATCGGTAAATGGCA-3′ (reverse); *Sdha*: 5′-TATGGTGCAGAAGCTCGGAAG-3′ (forward), 5′-ACTCATCGACCCGCACTTTG-3′ (reverse); *Ppargc1b*: 5′-CAGTACAGCCCCGATGACTC (forward), 5′-TTCGTAAGCGCAGCCAAGA-3′ (reverse); *Ppia*: 5′-AGCATACAGGTCCTGGCATC-3′ (forward), 5′- TTCACCTTCCCAAAGACCAC -3′ (reverse). *Ppia* gene was used as the internal control for normalization. The mRNA levels of genes of interest were normalized to the average levels of *Ppia*, and relative expression was calculated as the ratio to control (defined as 1).

RNA-seq and data processing

BMM from fCtrl and f*Kdm5c*$^{\Delta Vav}$ ($n = 3$) mice were seeded into 12-well plates followed by RANKL treatment for 0 hours (untreated), 16 hours, or 32 hours. Total RNA was extracted using the same method as described in qRT-PCR and sequenced by the Genomics Core at Van Andel Research Institute. Raw reads of RNA-seq data were mapped to *Mus musculus* (mm10). Subjunc v1.6.4 and featureCounts v1.6.4 were used to estimate the read counts on transcripts/gene exons. EdgeR v3.32.1 was used to identify differentially expressed genes. Genes with false discovery rate ≤ 0.05 (Benjamini-Hochberg–adjusted *P* values) were annotated as a differentially expressed gene between fCtrl and f*Kdm5c*$^{\Delta Vav}$ BMM. R package enrichR v3.1 was used to identify gene sets (Gene Ontology Biology Process 2021) enriched in the differentially expressed genes. Terms with $P < 0.05$ were considered significant. The top five terms ranked by combined score of up- and down-regulated genes are listed in Fig. 4A. Combined score is computed using the logarithm of Fisher exact test *P* value multiplex by *z* score.

Western blots

Cultured BMM were lysed in CHAPS buffer (Thermo Fisher Scientific) containing protein inhibitors. Ten to 40 μg of protein per sample was loaded to 10% SDS–polyacrylamide gel electrophoresis gels followed by transfer to polyvinylidene difluoride membranes. The membranes were then blocked with 5% nonfat milk in 1× tris-buffered saline plus 0.05% Tween 20 for 2 hours at room temperature and incubated with corresponding primary antibodies. Primary antibodies are as follows: anti-KDM5C (A301-034A, Bethyl Laboratories) or anti–β-actin (4967, Cell Signaling Technology) and OXPHOS rodent Western blots antibody cocktail (45-8099, Thermo Fisher Scientific). Blots were developed using the SuperSignalt West Dura Extended Duration Substrate (Thermo Fisher Scientific) and imaged using the ChemiDoc MP Imaging System (Bio-Rad).

Extracellular flux assay

Seahorse assay and ATP calculations were performed as described (*37*, *40*, *41*). OCR and ECAR of cultured BMM were measured using the Seahorse XF96 Extracellular Flux Analyzer. Briefly, 20,000 BMM were seeded to XF96 plates and treated with RANKL to induce osteoclastogenesis. Three days after RANKL treatment, cellular bioenergetics were assessed using the Seahorse XF Cell Mito Stress Test Kit with sequential addition of 1.5 μM oligomycin (inhibiting ATP synthesis), 3 μM carbonyl cyanide 4-(trifluoromethoxy) phenylhydrazone (uncoupling), 0.5 μM rotenone/antimycin A (inhibiting respiratory chain complexes I and III, respectively), and 10 mM monensin (stimulating Na + pumps on plasma membrane). Data were normalized to number of cells.

Mitochondrial ETC inhibition

Mitochondrial ETC complex I inhibitor piericidin A (#15379) and complex II inhibitor atpenin A5 (#11898) were purchased from Cayman Chemical, while complex III inhibitor myxothiazol (T5580) was purchased from MilliporeSigma. Cultured BMM were treated with DMSO (control group), piericidin A (0.2 μM), atpenin A5 (0.4 μM), myxothiazol (0.2 μM), or antimycin (0.2 μM) along with methyl-pyruvate (1 mM) and uridine (400 μM) followed by RANKL stimulation. All concentrations were referred to published papers (*42*, *43*) and optimized for BMM treatment. TRAP staining was performed when mature OCs appeared in the control group.

BMM retroviral transfection

BMM were transduced with control (pMSCV-ires-Thy1.1; pMIT) or PGC-1β–expressing (pMiT- *Pgc1*-β) retrovirus, as previously described (*44*, *45*). Briefly, 293T cells were transfected with vectors and Lipofectamine 2000 to generate retroviral supernatants. BMM were cultured in six-well plates with complete BMM media until 70% confluency was reached. Cells were then transfected with pMiT or pMiT-*Pgc1*-β retrovirus containing supernatants by centrifuging under 2500 rpm for 60 min at 30°C before being cultured in fresh complete BMM media. Two days after transfection, cells were digested with 0.5 mM EDTA, stained with phycoerythrin (PE)–Thy-1.1 antibody (12-0900-81, Thermo Fisher Scientific), and sorted using anti-PE microbeads (Miltenyi Biotec). Sorted BMM were seeded in 96-well plates for osteoclastogenesis, in 12-well plates for RNA extraction and Western blots, or in XF96 plates for Seahorse assays.

KDM5 inhibitor treatment

The KDM5i, KDM5A-IN-1, was purchased from MedChemExpress (catalog no. HY-100014). The half maximal effective concentration was defined as concentration at which the osteoclast numbers and area are halved compared to the control group. Cultured BMM from fCtrl mice or human monocytes were thereafter treated with DMSO or KDM5i for osteoclastogenesis assay, Seahorse assay, or RNA extraction.

Statistical analysis

All data were analyzed using GraphPad Prism software (version 9). An unpaired or paired Student's *t* test was performed for experiments with two groups for statistical significance calculation. A one-way or two-way analysis of variance (ANOVA) was performed to determine statistical significance between multiple groups. Actual *P* values are shown on each graph.

Supplementary Materials

This PDF file includes:

Figs. S1 to S8

REFERENCES AND NOTES

1. O. Lofman, L. Larsson, G. Toss, Bone mineral density in diagnosis of Osteoporosis: Reference population, definition of peak bone mass, and measured site determine prevalence. *J. Clin. Densitom.* **3**, 177–186 (2000).
2. S. Khosla, D. G. Monroe, Regulation of bone metabolism by sex steroids. *Cold Spring Harb. Perspect. Med.* **8**, a031211 (2018).
3. S. H. Mun, S. Jastrzebski, J. Kalinowski, S. Zeng, B. Oh, S. Bae, G. Eugenia, N. M. Khan, H. Drissi, P. Zhou, B. Shin, S. K. Lee, J. Lorenzo, K. H. Park-Min, Sexual dimorphism in differentiating osteoclast precursors demonstrates enhanced inflammatory pathway activation in female cells. *J. Bone Miner. Res.* **36**, 1104–1116 (2021).
4. D. N. Paglia, X. Yang, J. Kalinowski, S. Jastrzebski, H. Drissi, J. Lorenzo, Runx1 regulates myeloid precursor differentiation into osteoclasts without affecting differentiation into antigen presenting or phagocytic cells in both males and females. *Endocrinology* **157**, 3058–3069 (2016).
5. K. Astleford, E. Campbell, A. Norton, K. C. Mansky, Epigenetic regulators involved in osteoclast differentiation. *Int. J. Mol. Sci.* **21**, 7080 (2020).
6. F. Xu, W. Li, X. Yang, L. Na, L. Chen, G. Liu, The roles of epigenetics regulation in bone metabolism and osteoporosis. *Front. Cell Dev. Biol.* **8**, 619301 (2020).
7. P. A. Lingenfelter, D. A. Adler, D. Poslinski, S. Thomas, R. W. Elliott, V. M. Chapman, C. M. Disteche, Escape from X inactivation of *Smcx* is preceded by silencing during mouse development. *Nat. Genet.* **18**, 212–213 (1998).
8. M. G. Lee, J. Norman, A. Shilatifard, R. Shiekhattar, Physical and functional association of a trimethyl H3K4 demethylase and Ring6a/MBLR, a polycomb-like protein. *Cell* **128**, 877–887 (2007).
9. M. Scandaglia, J. P. Lopez-Atalaya, A. Medrano-Fernandez, M. T. Lopez-Cascales, B. del Blanco, M. Lipinski, E. Benito, R. Olivares, S. Iwase, Y. Shi, A. Barco, Loss of Kdm5c causes spurious transcription and prevents the fine-tuning of activity-regulated enhancers in neurons. *Cell Rep.* **21**, 47–59 (2017).
10. N. S. Outchkourov, J. M. Muiño, K. Kaufmann, W. F. J. van IJcken, M. J. G. Koerkamp, D. van Leenen, P. de Graaf, F. C. P. Holstege, F. G. Grosveld, H. T. M. Timmers, Balancing of histone H3K4 methylation states by the Kdm5c/SMCX histone demethylase modulates promoter and enhancer function. *Cell Rep.* **3**, 1071–1079 (2013).
11. P. Doss, M. Umair, J. Baillargeon, R. Fazazi, N. Fudge, I. Akbar, A. P. Yeola, J. B. Williams, M. Leclercq, C. J. Beauparlant, P. Beauchemin, G. F. Ruda, M. Alpaugh, A. C. Anderson, P. E. Brennan, A. Droit, H. Lassmann, C. S. Moore, M. Rangachari, Male sex chromosomal complement exacerbates the pathogenicity of Th17 cells in a chronic model of central nervous system autoimmunity. *Cell Rep.* **34**, 108833 (2021).
12. A. Dunford, D. M. Weinstock, V. Savova, S. E. Schumacher, J. P. Cleary, A. Yoda, T. J. Sullivan, J. M. Hess, A. A. Gimelbrant, R. Beroukhim, M. S. Lawrence, G. Getz, A. A. Lane, Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. *Nat. Genet.* **49**, 10–16 (2017).
13. J. C. Link, C. B. Wiese, X. Chen, R. Avetisyan, E. Ronquillo, F. Ma, X. Guo, J. Yao, M. Allison, Y. D. I. Chen, J. I. Rotter, J. S. el -Sayed Moustafa, K. S. Small, S. Iwase, M. Pellegrini, L. Vergnes, A. P. Arnold, K. Reue, X chromosome dosage of histone demethylase KDM5C determines sex differences in adiposity. *J. Clin. Invest.* **130**, 5688–5702 (2020).
14. M. K. Samanta, S. Gayen, C. Harris, E. Maclary, Y. Murata-Nakamura, R. M. Malcore, R. S. Porter, P. M. Garay, C. N. Vallianatos, P. B. Samollow, S. Iwase, S. Kalantry, Activation of Xist by an evolutionarily conserved function of KDM5C demethylase. *Nat. Commun.* **13**, 2602 (2022).
15. S. Iwase, F. Lan, P. Bayliss, L. de la Torre-Ubieta, M. Huarte, H. H. Qi, J. R. Whetstine, A. Bonni, T. M. Roberts, Y. Shi, The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. *Cell* **128**, 1077–1088 (2007).
16. C. N. Vallianatos, C. Farrehi, M. J. Friez, M. Burmeister, C. E. Keegan, S. Iwase, Altered gene-regulatory function of KDM5C by a novel mutation associated with autism and intellectual disability. *Front. Mol. Neurosci.* **11**, 104 (2018).
17. S. Iwase, E. Brookes, S. Agarwal, A. I. Badeaux, H. Ito, C. N. Vallianatos, G. S. Tomassy, T. Kasza, G. Lin, A. Thompson, L. Gu, K. Y. Kwan, C. Chen, M. A. Sartor, B. Egan, J. Xu, Y. Shi, A mouse model of X-linked intellectual disability associated with impaired removal of histone methylation. *Cell Rep.* **14**, 1000–1009 (2016).
18. F. Abidi, L. Holloway, C. A. Moore, D. D. Weaver, R. J. Simensen, R. E. Stevenson, R. C. Rogers, C. E. Schwartz, Novel human pathological mutations. Gene symbol: JARID1C. Disease: Mental retardation, X-linked. *Hum. Genet.* **125**, 345 (2009).
19. F. E. Abidi, L. Holloway, C. A. Moore, D. D. Weaver, R. J. Simensen, R. E. Stevenson, R. C. Rogers, C. E. Schwartz, Mutations in JARID1C are associated with X-linked mental retardation, short stature and hyperreflexia. *J. Med. Genet.* **45**, 787–793 (2008).
20. M. N. Weitzmann, T-cells and B-cells in osteoporosis. *Curr. Opin. Endocrinol. Diabetes Obes.* **21**, 461–467 (2014).
21. H.-F. Shen, W.-J. Zhang, Y. Huang, Y.-h. He, G.-s. Hu, L. Wang, B.-l. Peng, J. Yi, T.-t. Li, R. Rong, X.-y. Chen, J.-y. Liu, W.-j. Li, K. Ohgi, S.-W. Li, M. G. Rosenfeld, W. Liu, The dual function of KDM5C in both gene transcriptional activation and repression promotes breast cancer cell growth and tumorigenesis. *Adv. Sci.* **8**, 2004635 (2021).
22. S. Lemma, M. Sboarina, P. E. Porporato, N. Zini, P. Sonveaux, G. di Pompo, N. Baldini, S. Avnet, Energy metabolism in osteoclast formation and activity. *Int. J. Biochem. Cell Biol.* **79**, 168–180 (2016).
23. B. Li, W. C. Lee, C. Song, L. Ye, E. D. Abel, F. Long, Both aerobic glycolysis and mitochondrial respiration are required for osteoclast differentiation. *FASEB J.* **34**, 11058–11067 (2020).
24. H. N. Kim, F. Ponte, I. Nookaew, S. Ucer Ozgurel, A. Marques-Carvalho, S. Iyer, A. Warren, N. Aykin-Burns, K. Krager, V. A. Sardao, L. Han, R. de Cabo, H. Zhao, R. L. Jilka, S. C. Manolagas, M. Almeida, Estrogens decrease osteoclast number by attenuating mitochondria oxidative phosphorylation and ATP production in early osteoclast precursors. *Sci. Rep.* **10**, 11933 (2020).
25. F. Bost, L. Kaminski, The metabolic modulator PGC-1α in cancer. *Am. J. Cancer Res.* **9**, 198–211 (2019).
26. K. A. Ishii, T. Fumoto, K. Iwai, S. Takeshita, M. Ito, N. Shimohata, H. Aburatani, S. Taketani, C. J. Lelliott, A. Vidal-Puig, K. Ikeda, Coordination of PGC-1beta and iron uptake in mitochondrial biogenesis and osteoclast activation. *Nat. Med.* **15**, 259–266 (2009).
27. W. Wei, X. Wang, M. Yang, L. C. Smith, P. C. Dechow, J. Sonoda, R. M. Evans, Y. Wan, PGC1β mediates PPARγ activation of osteoclastogenesis and rosiglitazone-induced bone loss. *Cell Metab.* **11**, 503–516 (2010).
28. Y. Zhang, N. Rohatgi, D. J. Veis, J. Schilling, S. L. Teitelbaum, W. Zou, PGC1β organizes the osteoclast cytoskeleton by mitochondrial biogenesis and activation. *J. Bone Miner. Res.* **33**, 1114–1125 (2018).
29. J. Heward, L. Koniali, A. D'Avola, K. Close, A. Yeomans, M. Philpott, J. Dunford, T. Rahim, A. F. al Seraihi, J. Wang, K. Korfi, S. Araf, S. Iqbal, F. Bewicke-Copley, E. Kumar, D. Barisic, M. Calaminici, A. Clear, J. Gribben, P. Johnson, R. Neve, P. Cutillas, J. Okosun, U. Oppermann, A. Melnick, G. Packham, J. Fitzgibbon, KDM5 inhibition offers a novel therapeutic strategy for the treatment of KMT2D mutant lymphomas. *Blood* **138**, 370–381 (2021).
30. L. Wu, J. Cao, W. L. Cai, S. M. Lang, J. R. Horton, D. J. Jansen, Z. Z. Liu, J. F. Chen, M. Zhang, B. T. Mott, K. Pohida, G. Rai, S. C. Kales, M. J. Henderson, X. Hu, A. Jadhav, D. J. Maloney, A. Simeonov, S. Zhu, A. Iwasaki, M. D. Hall, X. Cheng, G. S. Shadel, Q. Yan, KDM5 histone demethylases repress immune response via suppression of STING. *PLoS Biol.* **16**, e2006134 (2018).
31. X. Liu, J. Secombe, The histone demethylase KDM5 activates gene expression by recognizing chromatin context through its PHD reader motif. *Cell Rep.* **13**, 2219–2231 (2015).
32. D. Grafodatskaya, B. H. Y. Chung, D. T. Butcher, A. L. Turinsky, S. J. Goodman, S. Choufani, Y. A. Chen, Y. Lou, C. Zhao, R. Rajendram, F. E. Abidi, C. Skinner, J. Stavropoulos, C. A. Bondy, J. Hamilton, S. Wodak, S. W. Scherer, C. E. Schwartz, R. Weksberg, Multilocus loss of DNA methylation in individuals with mutations in the histone H3 lysine 4 demethylase KDM5C. *BMC Med. Genomics* **6**, 1 (2013).

33. X. R. Bustelo, S. D. Rubin, K. L. Suen, D. Carrasco, M. Barbacid, Developmental expression of the vav protooncogene. *Cell Growth Differ.* **4**, 297–308 (1993).
34. B. E. Clausen, C. Burkhardt, W. Reith, R. Renkawitz, I. Forster, Conditional gene targeting in macrophages and granulocytes using LysMcre mice. *Transgenic Res.* **8**, 265–277 (1999).
35. S. L. Dallas, Y. Xie, L. A. Shiflett, Y. Ueki, Mouse Cre models for the study of bone diseases. *Curr. Osteoporos. Rep.* **16**, 466–477 (2018).
36. D. Csete, E. Simon, A. Alatshan, P. Aradi, C. Dobó-Nagy, Z. Jakus, S. Benkő, D. S. Győri, A. Mócsai, Hematopoietic or osteoclast-specific deletion of syk leads to increased bone mass in experimental mice. *Front. Immunol.* **10**, 937 (2019).
37. H. Guak, R. D. Sheldon, I. Beddows, A. V. Ark, M. J. Weiland, H. Shen, R. G. Jones, J. St-Pierre, E. H. Ma, C. M. Krawczyk, PGC-1beta maintains mitochondrial metabolism and restrains inflammatory gene expression. *Sci. Rep.* **12**, 16028 (2022).
38. D. Lu, J. Li, H. Liu, G. E. Foxa, K. Weaver, J. Li, B. O. Williams, T. Yang, LRP1 suppresses bone resorption in mice by inhibiting the RANKL-stimulated NF-kappaB and p38 pathways during osteoclastogenesis. *J. Bone Miner. Res.* **33**, 1773–1784 (2018).
39. J. Li, D. Lu, H. Dou, H. Liu, K. Weaver, W. Wang, J. Li, E. T. H. Yeh, B. O. Williams, L. Zheng, T. Yang, Desumoylase SENP6 maintains osteochondroprogenitor homeostasis by suppressing the p53 pathway. *Nat. Commun.* **9**, 143 (2018).
40. S. A. Mookerjee, A. A. Gerencser, D. G. Nicholls, M. D. Brand, Quantifying intracellular rates of glycolytic and oxidative ATP production and consumption using extracellular flux measurements. *J. Biol. Chem.* **292**, 7189–7207 (2017).
41. E. H. Ma, M. J. Verway, R. M. Johnson, D. G. Roy, M. Steadman, S. Hayes, K. S. Williams, R. D. Sheldon, B. Samborska, P. A. Kosinski, H. Kim, T. Griss, B. Faubert, S. A. Condotta, C. M. Krawczyk, R. J. DeBerardinis, K. M. Stewart, M. J. Richer, V. Chubukov, T. P. Roddy, R. G. Jones, Metabolic profiling using stable isotope tracing reveals distinct patterns of glucose utilization by physiologically activated $CD8^+$ T Cells. *Immunity* **51**, 856–870.e5 (2019).
42. H. Kong, C. R. Reczek, G. S. McElroy, E. M. Steinert, T. Wang, D. M. Sabatini, N. S. Chandel, Metabolic determinants of cellular fitness dependent on mitochondrial reactive oxygen species. *Sci. Adv.* **6**, eabb7272 (2020).
43. K. V. Tormos, E. Anso, R. B. Hamanaka, J. Eisenbart, J. Joseph, B. Kalyanaraman, N. S. Chandel, Mitochondrial complex III ROS regulate adipocyte differentiation. *Cell Metab.* **14**, 537–544 (2011).
44. B. Cordeiro, P. Jeon, G. M. Boukhaled, M. Corrado, O. Lapohos, D. G. Roy, K. Williams, R. G. Jones, S. Gruenheid, S. M. Sagan, C. M. Krawczyk, MicroRNA-9 fine-tunes dendritic cell function by suppressing negative regulators in a cell-type-specific manner. *Cell Rep.* **31**, 107585 (2020).
45. G. M. Boukhaled, B. Cordeiro, G. Deblois, V. Dimitrov, S. D. Bailey, T. Holowka, A. Domi, H. Guak, H. C. Chiu, B. Everts, E. J. Pearce, M. Lupien, J. H. White, C. M. Krawczyk, The transcriptional repressor polycomb group factor 6, PCGF6, negatively regulates dendritic cell activation and promotes quiescence. *Cell Rep.* **16**, 1829–1837 (2016).

Acknowledgments: We thank the current and former members of C. M. Krawczyk and T. Yang laboratories and S. Nowinski, R. Jones, and B. Williams for thoughtful discussions, and S. Craig and C. Anderson for editing. We are indebted to A. Shen from Calvin University for providing human PBMCs. This work was supported by the excellent staff in the Vivarium and Transgenics, Genomics, Pathology and Biorepository, and Metabolomics and Bioenergetics Cores at Van Andel Institute. **Funding:** This work was supported by the Van Andel Institute (VAI) Employee Impact Fund for Research awarded to T.Y. and C.M.K. (2021), VAI startup funds (C.M.K. and T.Y.), and the NIH (T.Y., R01AG061086). **Author contributions:** H.L., L.Z., T.Y., and C.M.K. conceived this study and designed experiments. H.L., L.Z., D.L., Y.L., and A.V.A. executed experiments and performed data analysis. H.L., L.Z., Y.L., and D.L. generated figures. H.L. and L.Z. wrote the manuscript, and T.Y. and C.M.K. edited the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 30 November 2022
Accepted 6 March 2023
Published 5 April 2023
10.1126/sciadv.adg0731

Scientists ♥ Grand Rapids

The city that has something for everyone

A vibrant biomedical corridor. Diverse and extensive arts and entertainment. White sand beaches and deep forest trails a short drive away.

There's a lot to love—and discover—about Grand Rapids, Michigan. With about 1 million people across its metro area, the state's second largest city has for the past decade experienced an economic and population boom, signaling to the world that something special is found here.

It's why Van Andel Institute's founding family resisted early calls to establish the Institute in a larger, more well-known place outside of Grand Rapids. They knew the city was the right location for VAI to thrive.

Founded in 1996 through a generous endowment by Jay and Betty Van Andel, VAI has steadily grown under the leadership of Chairman and CEO David Van Andel to become a major Midwestern hub for leading-edge biomedical research. The Institute spurred the emergence of the Grand Rapids Medical Mile, a district that includes several major hospitals and Michigan State University College of Human Medicine.

It takes more than good career opportunities to create a welcoming home, however—and Grand Rapids has a great deal to deliver. "Grand Rapids has so many things that big cities offer, but the pace and quality of life are much better," says VAI investigator Yvonne Fondufe-Mittendorf, noting that she can readily get from her home to her lab at VAI—or pretty much anywhere else in Grand Rapids—within 15 minutes or so.

VAI's Chief Scientific Officer Peter Jones says that the city's appeal is particularly strong for faculty and staff with children, who have access to affordable homes in convenient locations. "This is a marvelous place to bring up a family," he says. In fact, in 2023, *US News & World Report* ranked Grand Rapids #20 on its roster of best places to live in the US, with housing costs below the national average and an excellent school system—including City High School, which was ranked #1 in the state and #18 nationally in 2022.

Living in a smaller city doesn't necessarily require sacrifices in terms of access to entertainment and culture. Van Andel Arena brings headliners to town and houses minor league hockey and basketball teams. In addition to major arts hubs like the Grand Rapids Art Museum and the DeVos Performance Hall, the city also supports an extensive network of local artists and venues. "My youngest son loves theater, and we've gone to see the Broadway shows that come through," say VAI investigator Russell Jones. "We've already seen 'Hamilton' twice."

And in 2012, Grand Rapids was crowned "Beer City USA" in recognition of the more than 40 breweries in the area. But there's more than just beer on tap, as the city is also home to multiple craft distillers. Oenophiles can enjoy Michigan's wine region near Traverse City and Old Mission Peninsula just two hours north.

Nearby, Lake Michigan is home to a number of beautiful waterfront spots within easy driving distance from the city. There are numerous other inland lakes nearby, as well as myriad bike and hiking trails for outdoors enthusiasts. Furthermore, the 4,800-mile North Country Trail, which spans eight states, runs right past Grand Rapids. For those craving a trip to the big city, Detroit and Chicago are both only a few hours away. And for longer distance excursions—either for work or fun—Gerald R. Ford International Airport is close by and has direct services to more than 30 North American cities.

Grand Rapids only continues to expand. A major redevelopment of the downtown riverfront to include more green space and an open-air music venue is in motion. New hotels, restaurants, and entertainment venues are continually springing up.

Given this wealth of offerings and amenities, Fondufe-Mittendorf has quickly become a strong advocate for Grand Rapids. "It's beautiful," she says. "I love it."

VAN ANDEL INSTITUTE®
INNOVATION UNLEASHED

vai.org

YOUR NEXT STEP IS HERE

Now recruiting assistant, associate and full professors for all research areas:

EPIGENETICS
NEURODEGENERATION
METABOLISM & NUTRITION
CELL BIOLOGY
STRUCTURAL BIOLOGY

vai.org
616-234-5000