USC Michelson Center for Convergent Bioscience
January 2023 Quarterly Report

Contents

Executive Summary

We celebrate the 5-year anniversary of the inauguration of the USC Michelson Center for Convergent Bioscience and the 8-year anniversary of the construction groundbreaking of Michelson Hall, home to the Center. Thanks to a generous $50 million gift from retired orthopedic spinal surgeon Gary Michelson and his wife, Alya Michelson, the $185 million, 190,000-sf building provides the infrastructure that enables research leading to the development of life-saving therapeutics and devices, addressing the most critical health issues of our times.

Michelson Faculty — USC researchers whose laboratory is located at Michelson Hall and/or who are directors of Michelson Hall centers, institutes, and core facilities — continue to grow in number, currently 36 from the original 20 named faculty, and hail from three USC schools: the Viterbi School of Engineering, the Keck School of Medicine and the Dana and David Dornsife College of Letters, Arts and Sciences. Michelson Hall also houses five Centers/Institutes and five research core facilities, detailed later in this report.

Drawn from across a broad spectrum of research fields, ranging from computational sciences and applied mathematics to nanoengineering,

biology, and medicine, Michelson Faculty, along with hundreds of postdocs, students and research administrative staff, represent the human capital supporting the convergence of research at USC. Diverse in age, ethnicity and expertise, they interact and create new collaborative interfaces with researchers at USC and world-wide, driving technological innovation spanning the entire science, engineering and medical enterprise.

The Michelson Center January 2023 Quarterly Report presents a narrative of the Center's efforts and highlights the research being conducted by Michelson faculty during the period October-December, 2022. It presents the scientific and social impact, milestones and accolades achieved by its faculty, shared research facilities, institutes, and centers, as well as financial information on research labs and core facilities.

Since 2017, Michelson Faculty submitted an impressive 1160 proposals to external sponsors as principal investigators (PIs) and/or co-principal investigators (co-PIs), requesting almost $1.7 billion in funds; 320 proposals totaling over $234 million were awarded by federal and state agencies and government offices, corporations and foundations as grants, subcontracts or cooperative agreements, with anticipated obligated funding exceeding $304 million. Almost 40 grant proposals were submitted in 2022 alone

by Michelson Faculty as PIs or co-PIs, with over $1.4 million in awards made during the 2022 4th quarter.

New arrivals to Michelson Hall include Dr. Peter Foster, who joined USC in January 2023 as an Assistant Professor of Physics and Astronomy. As a member of Michelson's Bridge Institute, his research on the physics of living systems focuses on the cellular cytoskeleton and active matter physics. Also new to Michelson this January is Dr. Rehan Kapadia, Associate Professor of Electrical and Computer Engineering, who is serving as the new director of the John O'Brien Nanofabrication Laboratory core facility. Dr. Vsevolod Katritch, Associate Professor of Quantitative and Computational Biology and Chemistry, joined Dr. Charles McKenna, Professor of Chemistry and Pharmacology & Pharmaceutical Sciences, as co-director of the newly established Center for New Technologies in Drug Discovery and Development (CNT3D).

Recognition of research conducted by Michelson faculty is demonstrated by the continued number of national and international awards and accolades received.

- Dr. Carl Kesselman was named *2023 IEEE Fellow* by the Institute of Electrical and Electronics Engineers, a distinction reserved for select IEEE members whose extraordinary accomplishments in any of the IEEE fields of interest are deemed fitting for this prestigious grade elevation.

- Dr. Remo Rohs was elected *American Association for the Advancement of Science (AAAS) Fellow*, in recognition of his contributions to science and technology.

- Dr. Scott Fraser was one of three recipients of the Southern California Biomedical Council (SoCalBio) *2022 Innovation Awards.* Dr. Fraser was honored with the *2022 Outstanding Researcher Award*. The award recognizes those who "embody the spirit of innovation" and exemplify "the best that our region has to offer in entrepreneurship, research, and education."

- Dr. Stacey Finley received the *Cancer Warrior Award* from the University of Colorado Comprehensive Cancer Center in recognition of her "outstanding contribution and achievement in cancer research and devotion to training future cancer warriors."

- Dr. Steve Kay was designated a *2022 Clarivate Highly Cited Researcher*, one of 10 USC faculty recognized by Clarivate for the demonstrated, significant and broad influence of their research as reflected in their publication of multiple highly cited papers over the last decade.

- Dr. Ellis Meng was awarded a $11.7 million grant from the National Institutes of Health to create the Center for Autonomic Nerve Recording and Stimulation Systems (CARSS).

The featured article this quarter was published in the journal Nature — *Structure-based design of bitopic ligands for the μ-opioid receptor* — and reports on research by Dr. Vsevolod Katritch, in collaboration with researchers at the USC Bridge Institute, Stanford, Memorial Sloan Kettering, Scripps Research, and others nationally, demonstrating that chemically linking fentanyl to the sodium pockets within nerve cell receptors could block the harmful side effects of the synthetic opioid fentanyl and still reduce pain. The research addresses a critical societal need, given the opioid crises in the US, with nearly 70,000 opioid-related overdose deaths in 2020 alone, and holds promise that a less harmful version of fentanyl will work in humans, leading to safer pain-relieving drugs.

The highlighted articles include studies by CSI-Cancer Center director Dr. Peter Kuhn — *Defining a liquid biopsy profile of circulating tumor cells and oncosomes in metastatic colorectal cancer for clinical utility* — published in the journal Cancers (Basel). The research aims to validate the use of liquid biopsy as a prognostic tool to characterize pathophysiology and mechanisms of cancer metastasis.

Also highlighted is the article by Dr. Remo Rohs, *Physicochemical models of protein–DNA binding with standard and modified base pairs*, published

in the Proceedings of the National Academy of Sciences (PNAS). Dr. Rohs' research aims to elucidate the mechanisms by which DNA-binding proteins recognize genomic target sites and play important roles in various cellular processes.

We celebrate the scientific and social impacts and the advances in research achieved by Michelson Faculty, paving new frontiers in medicine and health. We celebrate the transformative potential of the Michelson Center for Convergent Bioscience to leverage the expertise of faculty, staff and students, realizing Dr. Michelson's vision to develop new life-saving devices and therapeutics, to reach "undreamed-of advances in the biological sciences in the near future" and ultimately "make life less unfair."

And it is with sincere and profound gratitude that we acknowledge that such achievements could not have been realized without the generous support of Dr. Gary Michelson and his wife Alya Michelson.

Happy Anniversary Michelson Center!
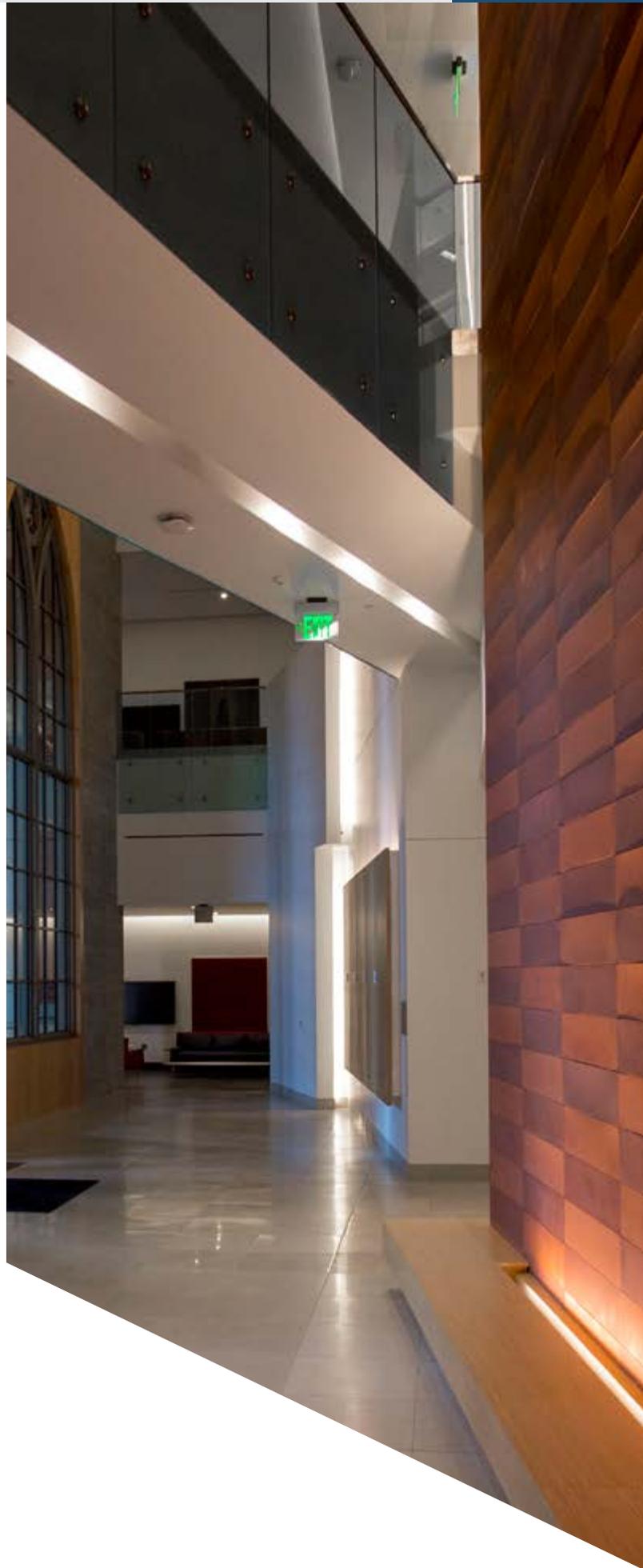
Five-Year Celebration

Creating the medical research infrastructure that drives a model for highly collaborative convergent projects and targeted science, enabling researchers to translate bold ideas into therapies, saving lives and empowering brilliant minds to heal the world.

Michelson Medical Research Foundation

Realizing a Transformative Vision

We celebrate the 5-year anniversary of the grand opening of the USC Michelson Center for Convergent Bioscience, thanks to a generous $50 million gift from retired orthopedic spinal surgeon Gary K. Michelson and his wife, Alya Michelson.

The $185 million, 190,000-sf building capped nearly seven years of planning and design, including three years of construction. Designed by HOK and erected by DPR Construction, the work involved a half dozen engineering firms and subcontractors.

It is the largest building on the University Park campus and was designed with an eye toward space flexibility to accommodate state-of-the instrumentation, infrastructure, conference rooms, Class 100 and 1,000 Nanotechnology cleanrooms, classrooms and lab space for Engineering and Biomedical Sciences, as well as scholarship in the Arts.

Michelson Hall is home to five core laboratories supporting research in imaging, biomolecular characterization, development of new peptides and antibody-like proteins, among many others. Its five Centers and Institutes are catalysts for innovation in drug discovery and development, informatics, imaging, medical and clinical the research. Thirty-six USC professors named as "Michelson Faculty" work across disciplines to form new, integrative research frameworks that allow faculty, postdocs, staff and students to work more efficiently and intuitively alongside each other.

Convergence: *from the Latin convergere, to "incline together"*

Grounded in Michelson Hall, research teams of the USC Michelson Center for Convergent Bioscience are tackling and pushing the boundaries of new frontiers in medicine and health. Interdisciplinary collaboration is leading to the development of new life-saving devices and therapeutics, responding to societal needs and delivering innovative solutions to address what the National Science Foundation calls complex and "vexing research problems."

"What you need to do is build runways, not fences. This building is a runway that brings together people who might never meet."

Steve Kay

From Blueprint to Building — Michelson Hall: By the Numbers

- October 23 2014: Groundbreaking of Michelson Hall construction

- > 275 construction workers and tradespeople spent a combined 699,700 hours to build Michelson Hall

- > 50,000 yards of soil exported

- 4.5 million pounds of concrete and rebar poured-in-place floors and walls

- 2,000-plus steel embeds used for a suspended clean room plenum steel and exterior masonry

- November 2017: Inauguration of the USC Michelson Center for Convergent Bioscience

Michelson Timeline

Groundbreaking date:
October 23, 2014

Andrea Armani
Carl Kesselman
James Hicks
Jeremy Mason
Niema Pahlevan
Peter Kuhn
Remo Rohs
Scott Fraser
Vsevolod Katritch

Agilent Center of
Excellence in Biomolecular
Characterization
Ellis Meng
Eunji Chung
Richard Roberts
Stacey Finley
Steve Kay
Terry Takahashi
Vadim Cherezov

2014   **2017**   **2017**   **2017**   **2018**

Michelson Hall inaugurated: December 2017

Center for Peptide and
Protein Engineering

Center for Discovery
Informatics

Convergent Science
Institute in Cancer

The Bridge Institute

Translational Imagining
Center

*"It's an interesting quality of science.
Once you solve the problem, the
solution seems almost obvious."*

Gary Michelson

Below you can find the dates researchers, centers and cores began their association with Michelson.

2019

Center of Excellence in Nano Imaging

Valery Fokin

2020

Cristina Zavaleta
Dani Byrd
Justin Haldar
Khalil Iskarous
Krishna Nayak
Shrikanth Narayanan

2021

Charles McKenna
Cornelius Gati
Gianluca Lazzi
James Boedicker
Jayakanth Ravichandran
Kate White
Mercedeh Khajavikhan
Moh El-Naggar

Dynamic Imaging Science Center

John O'Brien Nanofabrication Laboratory

2021

2022

Center for New Technologies in Drug Discovery & Development

Jennifer Treweek

Maryam Shanechi

Yasser Khan

2023

Peter Foster
Rehan Kapadia

"We try to be involved in areas where nobody else wants to go. Where it is hard. We try to be innovative in philanthropy and lead with the principles that Dr. Michelson applied to all his devices and discoveries: Be innovative. Don't be afraid to disrupt things, and be creative."

Alya Michelson

Fundraising & Financials

Data reflect a reporting period from October 1 to December 31, 2022 of grants submitted by/awarded to Michelson faculty.

Proposals Awarded

	Grant proposals awarded as PI		Grant proposals awarded as Co-PI	
	# of proposals	Total obligated	# of proposals	Total obligated
Andrea Armani	1	$300,691.00		
Cristina Zavaleta	1	$299,998.00		
James Boedicker			1	$300,691.00
James Hicks			1	$225,000.00
Niema Pahlevan	1	$60,478.00		
Peter Kuhn	1	$225,000.00		
Shrikanth Narayanan			1	$49,866.00
Total	*4*	*$886,167.00*	*3*	*$575,557.00*

Proposals Submitted

	Grant proposals submitted as PI		Grant proposals submitted as Co-PI	
	# of proposals	Total requested	# of proposals	Total requested
Carl Kesselman	1	$749,925.00		
Charles Mc Kenna	2	$3,237,670.00		
Cornelius Gati	2	$5,876,475.00		
Dani Byrd			1	$1,200,000.00
Eun Ji Chung	2	$1,033,324.00		
Gianluca Lazzi	3	$6,973,598.00		
Jayakanth Ravichandran	1	$450,000.00		
Jennifer Treweek	1	$100,000.00		
Kate White	2	$294,000.00		
Krishna Nayak			1	$1,200,000.00
Mercedeh Khajavikhan	1		1	$750,000.00
Moh El-Naggar	1	$1,600,000.00		
Niema Pahlevan	2	$243,398.00		
Peter Kuhn	5	$3,853,749.00		
Rehan Kapadia	1	$265,580.00	1	
Scott Fraser			2	$192,844.00
Shrikanth Narayanan	1	$1,200,000.00		
Stacey Finley	1	$11,034,993.00		
Steve Kay	1	$441,750.00		
Terry Takahashi			1	$197,352.00
Vadim Cherezov	1	$89,566.00	1	$618,750.00
Vsevolod Katritch	1	$618,750.00	1	$500,000.00
Yasser Khan	1	$618,928.00		
Total	*30*	*$38,681,706.00*	*9*	*$4,658,946.00*

Number of grants awarded (by grant type) As PI
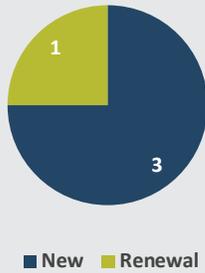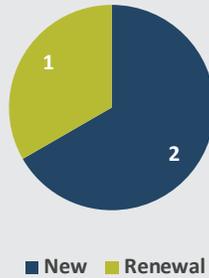
- New
- Renewal

Number of grants awarded (by grant type) As Co-PI

- New
- Renewal

Number of grants awarded (by sponsor type) As PI

- Federal
- Foundation

Number of grants awarded (by sponsor type) As Co-PI

- Federal
- Foundation

Number of grants submitted (by grant type) As PI

- New
- Continuation
- Resubmission
- Renewal
- Preproposal
- Revisions

Number of grants submitted (by grant type) As Co-PI

- New
- Continuation
- Preproposal

Number of grants submitted (by sponsor type) As PI

- Federal
- Institution of Higher Education - Private
- Foundation
- Industry
- Institution of Higher Education - State
- Hospital

Number of grants submitted (by sponsor type) As Co-PI

- Federal
- Foundation
- Industry
- Institution of Higher Education - State
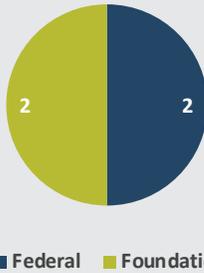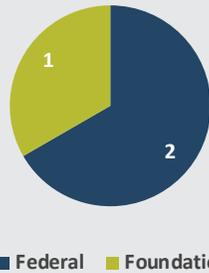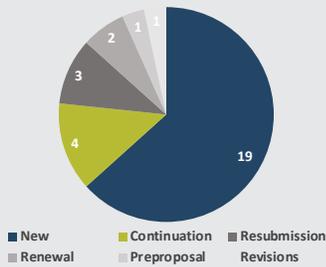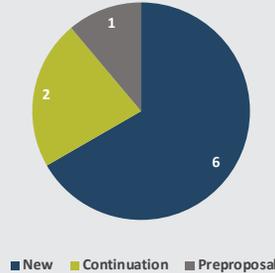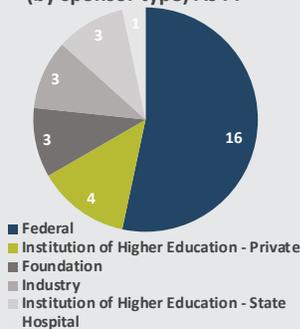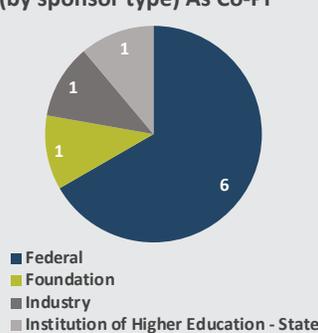
Research Advancement Update

Ongoing conversations are being held with Amgen around the Cryo E/M instruments and expanding the partnership to include another such instrument to be donated to USC.

Meet Our Faculty

Faculty and research teams are forging new frontiers in biomedical research, leading to life-saving therapies.

MICHELSON FACULTY

Andrea Armani, Ph.D.
USC Viterbi School of Engineering

Ray Irani Chair in Chemical Engineering and Materials Science and Professor of Chemical Engineering and Materials Science, Biomedical Engineering, Electrical Engineering-Electrophysics, and Chemistry

James Boedicker, Ph.D.
USC Dornsife College of Letters, Arts, and Sciences

Associate Professor of Physics and Astronomy and Biological Sciences

Dani Byrd, Ph.D.
USC Dornsife College of Letters, Arts, and Sciences

Professor of Linguistics

Vadim Cherezov, Ph.D.
USC Dornsife College of Letters, Arts, and Sciences

Professor of Chemistry, Physics & Astronomy

Eun Ji Chung, Ph.D.
USC Viterbi School of Engineering

The Dr. Karl Jacob Jr. and Karl Jacob III Early Career Chair and Assistant Professor of Biomedical Engineering, Chemical Engineering and Materials Science, Surgery, and Medicine

Moh El-Naggar, Ph.D.
USC Dornsife College of Letters, Arts, and Sciences

Dean's Professor of Physics and Astronomy, and Professor of Physics, Biological Sciences, and Chemistry

Diverse in age, ethnicity and expertise, and drawn from across a broad spectrum of research fields, ranging from computational sciences and applied mathematics to nanoengineering and biology, Michelson faculty create new collaborative interfaces and technological innovation spanning the entire science, engineering and medical enterprise.

Stacey Finley, Ph.D.
USC Viterbi School of Engineering

Gordon S. Marshall Early Career Chair and Associate Professor of Biomedical Engineering and Biological Sciences

Valery Fokin, Ph.D.
USC Dornsife College of Letters, Arts, and Sciences

Professor of Chemistry

Peter Foster, Ph.D.
USC Dornsife College of Letters, Arts, and Sciences

Assistant Professor of Physics and Astronomy

Scott E. Fraser, Ph.D.
Provost's Office

Provost Professor of Biological Sciences, Biomedical Engineering, Physiology and Biophysics, Stem Cell Biology and Regenerative Medicine, Pediatrics, Radiology and Ophthalmology Elizabeth Garrett Chair in Convergent Bioscience; Director of Science Initiatives

Cornelius Gati, Ph.D.
USC Dornsife College of Letters, Arts, and Sciences

Assistant Professor of Biology

Justin Haldar, Ph.D.
USC Viterbi School of Engineering

Associate Professor of Electrical and Computer Engineering and Biomedical Engineering

James Hicks, Ph.D.
USC Dornsife College of
Letters, Arts, and Sciences

Research Professor, Biological Sciences

Khalil Iskarous, Ph.D.
USC Dornsife College of
Letters, Arts, and Sciences

Associate Professor of Linguistics

Rehan Kapadia, Ph.D.
USC Viterbi School of Engineering

Associate Professor of Electrical and
Computer Engineering

Vsevolod Katritch, Ph.D.
USC Dornsife College of
Letters, Arts, and Sciences

Assistant Professor of Biological Sciences
and Chemistry

Steve Kay, Ph.D., D.Sc.
Keck School of Medicine of USC

Director of Convergent Bioscience and
Provost Professor of Neurology, Biomedical
Engineering and Biological Sciences

Carl Kesselman, Ph.D.
USC Viterbi School of Engineering

Dean's Professor of Industrial and Systems
Engineering; Director, Center for Discovery
Informatics; Information Sciences Institute
Fellow

**Mercedeh Khajavikhan,
Ph.D.**
USC Viterbi School of Engineering

Professor of Electrical and Computer
Engineering

Yasser Khan, Ph.D.
USC Viterbi School of Engineering

Assistant Professor of Electrical and
Computer Engineering

Peter Kuhn, Ph.D.
USC Dornsife College of
Letters, Arts, and Sciences

Dean's Professor of Biological Sciences;
Professor of Medicine, Biomedical
Engineering, Aerospace & Mechanical
Engineering, and Urology

Gianluca Lazzi, Ph.D.
Keck School of Medicine of USC

Provost Professor of Ophthalmology, Electrical Engineering, Clinical Entrepreneurship and Biomedical Engineering; Fred H. Cole Professorship in Engineering; Director, Institute for Technology and Medical Systems (ITEMS); Director for Bioengineering Mentorship

Jeremy Mason, Ph.D.
Keck School of Medicine of USC

Assistant Professor of Research, Department of Urology

Charles E. McKenna, Ph.D.
USC Dornsife College of Letters, Arts, and Sciences

Professor of Chemistry & Pharmacology and Pharmaceutical Sciences

Ellis Meng, Ph.D.
USC Dornsife College of Letters, Arts, and Sciences

Dwight C. and Hildagarde E. Baum Chair in Biomedical Engineering

Shrikanth Narayanan, Ph.D.
USC Viterbi School of Engineering

University Professor and Niki & C. L. Max Nikias Chair in Engineering; Professor of Electrical and Computer Engineering, Computer Science, Linguistics, Psychology, Neuroscience, Otolaryngology-Head and Neck Surgery, and Pediatrics; Research Director, Information Science Institute; Director, Ming Hsieh Institute

Krishna S. Nayak, Ph.D.
USC Viterbi School of Engineering

Professor of Electrical and Computer Engineering, Biomedical Engineering, and Radiology; Director, Dynamic Imaging Science Center; Director, Signal and Image Processing Institute; Director, Magnetic Resonance Engineering Laboratory

Niema Pahlevan, Ph.D.
USC Viterbi School of Engineering

Assistant Professor of Aerospace and Mechanical Engineering

Jayakanth Ravichandran, Ph.D.
USC Viterbi School of Engineering

Associate Professor of Chemical Engineering and Materials Science

Richard W. Roberts, Ph.D.
USC Dornsife College of Letters, Arts, and Sciences

Professor and Chair, Mork Family Department of Chemical Engineering and Materials Science and Professor of Chemistry, Chemical Engineering, and Biomedical Engineering

Remo Rohs, Ph.D.
USC Dornsife College of
Letters, Arts, and Sciences

Professor of Biological Sciences, Chemistry,
Physics & Astronomy, and Computer Science

Maryam Shanechi, Ph.D.
USC Viterbi School of Engineering

Associate Professor of Electrical and
Computer Engineering

Terry Takahashi, Ph.D.
USC Dornsife College of
Letters, Arts, and Sciences

Assistant Professor (Research) of Chemistry

Jennifer Treweek, Ph.D.
USC Viterbi School of Engineering

Assistant Professor

Kate White, Ph.D.
USC Dornsife College of
Letters, Arts, and Sciences

Gabilan Assistant Professor of Chemistry;
Director, Pancreatic Beta Cell Consortium
(PBCC)

Cristina Zavaleta, Ph.D.
USC Viterbi School of Engineering

Assistant Professor of Biomedical
Engineering

Institutes & Centers

The goal of a research center is to enable interaction between faculty, scholars, students, and industry, to enhance research opportunities, academic excellence, real-world problem solving, and knowledge creation and dissemination.

V. Kumar. J. of the Acad. Mark. Sci. (2017) 45:453–458.

The University of Southern California is home to over 200 research centers and institutes, spanning a wide array of disciplines and specialties within and across campus, as well as off-campus locations, and serves as a locus of research for collaborative groups of investigators, driving the frontiers of knowledge forward.

The USC Michelson Center for Convergent Bioscience is hub to five such institutes and centers, providing the infrastructure and staff to support a diverse network of scientists, engineers and scholars from the USC Dornsife College of Letters, Arts and Sciences, USC Viterbi School of Engineering and Keck School of Medicine. The Centers' rich and interdisciplinary research ecosystem and forward-thinking initiatives aim to solve some of the greatest intractable problems of the 21st century – from cancer, to neurological disease, to cardiovascular disease — propelling medical advancements from the bench to the bedside at an unprecedented pace.

Research centers and institutes at Michelson Hall provide leadership and direction to programs designed to foster fundamental creative discoveries, innovative research strategies, and their applications, expand the knowledge base, and promote the highest level of scientific integrity, and social responsibility in the conduct of science. They leverage the expertise of USC faculty, fostering national and international collaborations and supporting Dr. Gary Michelson's vision of a "more open, convergent, and information-sharing approach" to research so that "researchers aren't just working harder – they are working smarter."

"There's no choice but to embrace the need for trans-disciplinary teams to tackle complex problems. With the tools and infrastructure here, we're becoming a destination for training the world about how convergence can be applied to medicine, science and engineering."

Steve Kay

Convergent Science Institute in Cancer (CSI-Cancer)

Institute Directors:
Dr. Peter Kuhn, Director, Dean's Professor of Biological Sciences and Professor of Medicine, Biomedical Engineering, and Aerospace & Mechanical Engineering and Urology
Dr. James Hicks, Deputy Director, Research Professor of Biological Sciences
Dr. Jeremy Mason, Director of Data Science, Assistant Professor of Research, Urology
Dr. Liya Xu, Director of Applied Genomics, Assistant Professor of Research, Ophthalmology, CHLA

USC Michelson Center CSI-Cancer aims to integrate patient, model system, and high-content single cell data to translate clinically observed correlations into a mechanistic understanding of the physical and biological underpinnings of cancer dynamics. The organizing framework of the physical dynamics of cancer at the lab focuses on the spatial distributions and temporal evolution of the disease at the cellular, human, and population scale.

Three complementary research projects, supported by one shared resource core, address the center's aims of:

- developing a quantitative description of the transition between limited and disseminated cancer;

- demarcating the complexity and alterations over time of a cancer in all its phases;

- describing the temporal evolution of disease in patients with poor prognosis vs favorable prognosis.

Founded by Dr. Peter Kuhn, each project is led by a team of patients, medical doctors, scientists and students focused on improving the outcomes for patients with cancer directly through earlier detection and better treatment management throughout a patient's lifetime. As such, the center aims to better understand how to evaluate individual patients' fitness for their next course of cancer treatment, using body sensors and cameras, as well as a cell phone app for patients to report symptoms, to quantify patient fitness and help predict the need for hospitalization or other urgent interventions.

CSI-Cancer has built a large collaborative interdisciplinary network with a number of schools and centers at USC (Viterbi, Dornsife, Keck, the USC Norris Comprehensive Cancer Center, the USC Roski Eye Institute and USC Stevens), as well as with federal, state, private and corporate agencies, institutes, hospitals and research facilities, nationally and internationally, to disseminate its capabilities and to collaborate on evolving physical sciences concepts that could address specific challenges as part of the organizing framework within the institute and the scientific community at large. A comprehensive list of collaborators can be found at the CSI website. Since its creation, CSI-Cancer researchers have published over 150 manuscripts. These papers have been cited over ten thousand times in that time period.

Accolades

Dr. Kuhn was among a select group of leaders in the cancer research community to be invited by President Biden and Dr. Biden to participate in a ceremony at the White House announcing the restart of the Cancer Moonshot initiative to decrease cancer death by 50% in the next 25 years. Since the initiative was first introduced in 2016, Dr. Kuhn and his team have participated in the BloodPAC Consortium to enhance collaboration and data sharing between academic and industry partners, a key goal of the Moonshot.

Educational Outreach

Educational outreach is an important component of the center's activities. In partnership with Neighborhood Academic Initiative (NAI), a number of CURE students from the REACH program led a five-session workshop at USC for 110 local high school students. The Cancer Undergraduate Research Experience (CURE) program is the undergraduate program of USC Michelson CSI-Cancer. Each student is assigned a mentor who helps them work on a specific project. CSI ACE is a comprehensive program designed to help students with an earned Associate's degree complete their academic journey to the Bachelor's degree on time, providing a range of financial, academic, and personal support, including individualized academic advising and career development support, as well as tuition, textbook, and transportation assistance.

The Bridge Institute

Center Directors:
Dr. Scott Fraser, Director, Provost Professor of Biological Sciences and Biomedical Engineering, Elizabeth Garrett Chair of Convergent Biosciences
Dr. Kate White, Associate Director, Gabilan Assistant Professor of Chemistry, Director of the Pancreatic Beta Cell Consortium
Angela (Angie) Walker, Assistant Director, Bridge Institute

Over the past century, the sciences, engineering and medicine have become progressively more specialized, each diverging into an array of sub-disciplines. While this approach has led to significant increases in understanding human health, it has also resulted in silos of expertise that run deep and often remain isolated, hindering progress. The Bridge Institute aims to disrupt conventional thinking — encouraging researchers to think outside of the box — converging knowledge to achieve radical progress in improving the human condition.

Building upon the university's Michelson Center for Convergent Bioscience Initiative, the Bridge Institute unites eminent professors across the university, including the USC Dornsife College of Letters, Arts and Sciences, the USC Viterbi School of Engineering and the Keck School of Medicine of USC, as well as animators and cinematographers in the USC School of Cinematic Arts and the USC Institute for Creative Technologies, and technology transfer experts in the USC Stevens Center for Innovation. It serves as a launchpad for projects drawing on disparate disciplines of science, engineering, medicine and the arts, leveraging the strengths of individual departments and schools by providing an umbrella that trains fearless

researchers, supports convergence and empowers collaboration.

The mission of the Bridge Institute is to explore the interplay of molecules, cells and tissues through inclusive Team-Science and Art-Science frameworks to create new mechanistic insights that impact biomedical research and support science training.

Its values are Teamwork, Imagination, Innovation, Learning, Teaching, and Convergence.

Accolades
Institute director Dr. Scott Fraser was one of three award recipients of the Southern California Biomedical Council (SoCalBio) — the trade association for bioscience and digital health professionals across greater Los Angeles –honoring bioscience innovators during the 2022 SoCalBio Award Gala in December. Dr. Fraser was the recipient of the *2022 Outstanding Researcher Award*.

Events
The Bridge Institute and the USC World Building media Lab hosted a playlist of their current experience *World in a Call*. The 2-day event, held last November at Michelson Hall, highlighted ongoing art-science collaborative efforts, which use

immersive technologies to make complex scientific concepts approachable for diabetes research and education.

The Bridge Institute also sponsors a "C-STEM Lunch and Learn" "What I talk about, when I talk about…" seminar series. Topics covered in October and November included Cryo-Electron Microscopy (with Htet Khant, Senior Scientist, Cryo-Electron Microscopy, Viterbi); Electron Microscopy (with Carolyn Marks, Bio-EM Senior Scientist) and Click Chemistry (with Dr. Valery Fokin, Professor of Chemistry).

Educational Outreach
The Bridge UnderGraduate Science (BUGS) Summer Research Program began as a pilot program in 2015 with participation of only four USC faculty members as mentors, nine undergraduate (BUGS) and seven High School (BUGS Jr) students. By the summer of 2022, the BUGS and BUGS Jr Programs had grown to achieve their highest levels of participation to date, with 54 USC faculty mentors (from seven separate schools) in the joint program, along with 46 undergrads and 51 high school students.

Over the course of the past 8 years, the BUGS/Jr Programs have helped shape the educational and professional trajectories of 185 undergraduates and 136 high school students. Many of the undergraduates have reported matriculating on to medical/graduate programs – three of them earning honors as Fulbright Scholars after their participation in the BUGS program. Likewise, many of the high school students have since graduated and moved forward to community college or 4-year university programs, including at least four from the 2022 class that are now studying at USC.

Center for Discovery Informatics (CDI)

Center Director:
Dr. Carl Kesselman, Director, Dean's Professor of Industrial and Systems Engineering and Professor of Industrial and Systems Engineering, Computer Science, and Preventive Medicine

Convergence creates the opportunity to solve the most intractable biomedical problems of our time, and discovery informatics provides the foundation to do it. In convergent bioscience, the ability to generate novel discoveries is now strongly dependent on our ability to find, organize, integrate and analyze increasingly large, complex and dynamic data.

By definition convergent bioscience research is interdisciplinary and data from those disciplines is heterogeneous, with multiple modalities including microscopy, mass spectrometry, flow cytometry, sequencing and other instrumentation. The Center provides the research infrastructure to tie these disparate threads of convergent bioscience research together. Its goal is to transform how knowledge is created, explored and translated into benefits for humankind.

CDI is building a shared research infrastructure, spanning engineering, computer science, social sciences and biomedical research, supporting and driving collaborating scientists to:

- Address the significant and challenging issue of how to better capture, organize and share data as part of the discovery process;

- Understand how intertwining technology, with the daily practice of science, can create opportunities to develop and integrate advanced analysis, data mining, visualization and interaction methods.

Challenges of Big Data Informatics
Recent studies have indicated that just managing and organizing data may consume the majority of a researcher's time (50-80%) and that scientists infrequently share data unless required to do so. According to Dr. Kesselman, without discovery informatics scientists are inefficient, have low success rates in experimentation, create irreproducible results (only 10% are reproduced effectively), and miss opportunity for collaboration and transformative discovery.

At CDI, discovery informatics plays an integral and critical role to address the challenges associated with data management by providing the fundamental abstractions, methods, tools and infrastructure to enable transformative, data-driven discovery in complex biological systems. CDI's broad-based, computer science and informatics research program combines elements of data-driven science, grid and cloud computer, user-centered design, information visualization, data integration, web technologies and service-oriented architecture, big-data analysis and analytics, data management and high performance computing to generate new approaches to research and create a unique research platform unique to the USC Michelson Center for Convergent Bioscience.
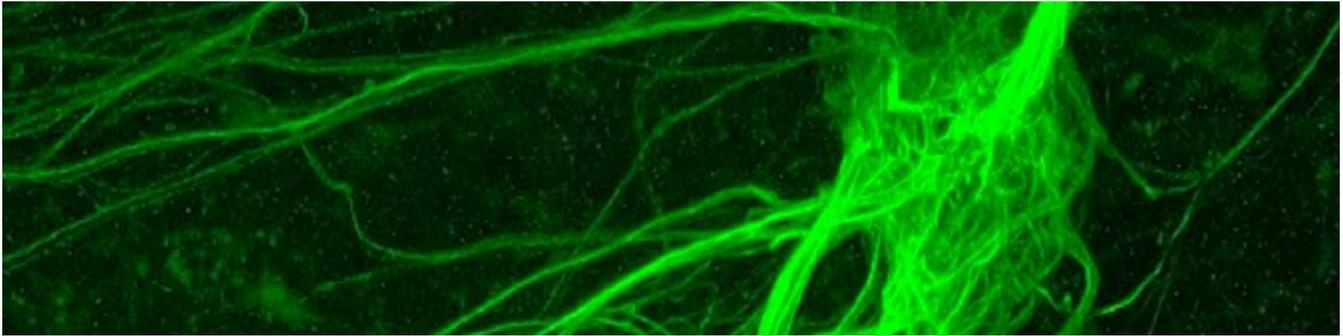
Translational Imaging Center (TIC)

Center Director:
Dr. Scott Fraser, Director, Director of Science Initiatives, Provost Professor of Biological Sciences and Biomedical Engineering, Elizabeth Garrett Chair of Convergent Biosciences

The Translational Imaging Center (TIC) serves dual, complementary roles as a research center and user facility and is led by Dr. Scott Fraser, who also serves as director of the Michelson Center's Bridge Institute. The TIC's mission is to develop new technologies for the imaging of biological structures and functions.

The center provides the research community with a unique mix of cutting-edge commercial multiphoton and confocal laser scanning microscopes, optical coherence tomography and Magnetic Resonance Imaging (MRI) microscopy systems, as well as custom-made Light Sheet systems. It is home to a diverse group of scientists, embracing the fields of Biology, Physics, Engineering, Chemistry, and Mathematics, creating a highly interdisciplinary environment, enhanced by scientific collaborations from the USC community and around the world.

Research supported by the TIC varies greatly and entertains questions ranging from embryonic development, to genetics, to neuroscience. In each of the TIC's research focuses there is a common theme: the use of advanced imaging tools to follow events as they take place inside an intact organism, bridging the gap between basic science research and science-based medicine and allowing it to spearhead key projects, giving rise to important biomedical devices and treatments in areas ranging from eye disease to cancer. Over 40 national and international patents/patent applications are associated to researchers at TIC.

Much of the work in the TIC is focused upon the study of tissue patterning in developing embryos, but recent efforts have moved the research towards models of disease as well. The diseases of primary interest have been Multiple Sclerosis and Alzheimer's Disease, with an increasing interest in Macular Degeneration and other eye diseases, with close collaborative efforts with USC's Roski eye Institute.

Awards
The Southern California Biomedical Council (SoCalBio) -- the trade association for bioscience and digital health professionals across Greater Los Angeles –honored three of the region's most esteemed bioscience innovators during the 2022 SoCalBio Award Gala in December. The Awards recognize and celebrate the contributions of Los Angeles Region bioscience entrepreneurs, researchers, and educators. The 2022 honorees were TIC director Dr. Fraser, recipient of the *Outstanding Researcher Award*; Dr. David Baltimore, President Emeritus; Judge Shirley Hufstedler Professor of Biology, Caltech, recipient of the *Life-Time Achievement Award*; and Dr. Marc Baum, President, Oaks Crest Institute, recipient of the *Outstanding Educator Award*.

Events

Michelson Hall serves as a venue for TIC's seminars and educational activities throughout the year, including its Visionary Seminar Series at USC that recently (November 2022) sponsored a talk by Dr. Peter Butler, Professor of Medicine and Director of the Larry Hillblom Islet Research Center at UCLA, titled *"Cellular dysfunction in diabetes and dementia, fighting to live another day."*
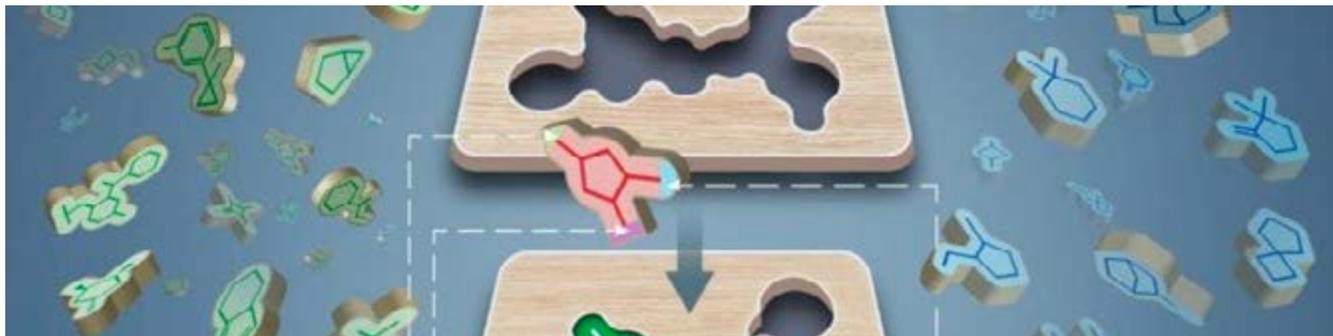
Center for New Technologies in Drug Discovery and Development (CNT3D)

Center Directors:
Dr. Vsevolod "Seva" Katritch, Associate Professor of Quantitative and Computational Biology and Chemistry
Dr. Charles McKenna, Professor of Chemistry and Pharmacology & Pharmaceutical Sciences

The Center for New Technologies in Drug Discovery and Development (CNT3D) was established as a research center at Dornsife College and joined Michelson Hall in 2022. The center is animated by a multidisciplinary group of researchers, who lead the development and application of cutting-edge approaches to address a wide range of targets and therapeutic indications.

The CNT3D is focused on applying breakthrough technologies in synthetic and medicinal chemistry, structural biology, computational chemistry, and machine learning/AI to drug discovery and development (3D). Its goal is to serve the USC biomedical community by facilitating cost-effective and well-curated entry points into drug discovery at the molecular level.

Research
The CNT3D research labs are developing several state-of-the-art platforms synergistically streamlining the drug discovery and development process.

One of the key platforms of the center includes V-SYNTHES, a giga-scale structure-based virtual screening technology for hit and lead discovery developed in Dr. Vsevolod Katritch's lab (Nature 2021). The technology has been validated on several therapeutic target classes, yielding high-quality lead series ready for further development.

The new V-SYNTHES 2.1, expanded with AI capabilities, is being tested in collaborations with USC labs and nationally, showing promising results for a wide range of clinical targets.

A new approach to expanding on-demand compound screening libraries using click chemistry has been developed by the CNT3D's Dr. Valery Fokin and has been validated in collaboration with the Katritch lab. Click chemistry is a class of chemical synthesis that uses simple, atom-economy reactions to join two molecular entities of choice. The research has yielded highly potent hits from the screening of 140 million compounds, which can be synthesized in-house.

Structure-based drug discovery, for example, by V-SYNTHES, is enabled by high-resolution structural information being generated by CNT3D researchers in the state-of-the-art facilities at Michelson Hall, in collaborative efforts with the Structural Biology Center @Bridge, which specializes in soluble proteins.

Other examples of ongoing research by CNT3D-associated faculty include: Dr. Vadim Cherezov, who has solved structures of >60 key therapeutic targets using advanced crystallography; the lab of Dr. Cornelius Gati, pushing the limits of cryo-EM technology in the new USC Dornsife Cryo-EM facility, and Dr Kate White's "molecule to

cell" mission that aims to optimize experimental methods for 3-dimensional visualization of single cells. The White lab works to characterize the cellular ultrastructure and develop integrative whole-cell modeling infrastructure to harmonize structural and mathematical representations of the cell across the scales of biology.

Novel prodrug technology is being developed in the McKenna lab (Gribble 2023, Antiviral Research featured article, 2022), and targets selective inhibition of DNA and RNA replication, yielding highly effective oral compounds against viral infections. The technology is widely adaptable to create more effective antivirals, including the CoV-2 virus. In addition, promising new compounds are being developed to combat cancer, fungus infections and bone diseases.

CNT3D will enable cost-effective identification of highly diverse, potent, bioavailable, target-selective, and drug-like ligands and lead series for any desired target, inviting USC researchers and the national and international research community to take advantage of its lead discovery pipeline to follow-up with preclinical and clinical development.

Events
The center will host a Drug Discovery Innovation Workshop, as a collaborative effort between CNT3D, USC's MESH academy and the Bridge Institute, with support from a pilot phase I grant from Dornsife College, to be held in Michelson Hall in February. The workshop will showcase novel advanced computational and experimental technologies for drug discovery being developed by research labs at USC and explore how these technologies can be most efficiently employed in drug discovery and development, bringing together the drug discovery community at USC and the greater LA area. The workshop will be comprised of a number of talks on specific 3D technologies and overall 3D efforts at USC as well as showcase a discussion panel on ways to bring together the USC 3D communities, with representatives from Dornsife College, as well as the Viterbi School of Engineering, the Mann School of Pharmacy and the Keck School of Medicine.

Accolades
A documentary created by German film-maker Ralph Loop explored novel hearing loss drug delivery research and featured CNT3D's Dr. McKenna; it was broadcast on ARTE (BBC of the EU) in French and German versions.

Core Facilities

Shared cores at Michelson Hall advance biomedical research by providing outstanding resources — including facilities, equipment, tools, databases, and services — that enable investigators to perform cutting-edge research.

Core Impact

Shared resource facilities, or cores, are central hubs of discovery, providing outstanding resources to the research community — equipment, tools, databases, repositories and expert services — enabling investigators to perform cutting-edge research. USC is home to over 60 core facilities (located on multiple campuses — University Park Campus, the Health Science Campus, Marina del Rey and other off-campus locations as well as Children's Hospital, Los Angeles), five of which are located at Michelson Hall.

Michelson Center cores offer a broad range of technical expertise, specialized consultation services and state-of-the-art instrumentation to researchers, including robotics, imaging, biomolecular characterization, development of new peptides and antibody-like proteins, among many others. They facilitate close collaboration between researchers from different fields at USC, as well as nationally and internationally.

Core facilities provide fee-for-service to researchers. They consolidate resources, increasing access to instrumentation and services that could otherwise be prohibitively expensive, and provide equitable access to niche and emerging areas of research.

In 2022, core facilities at Michelson Hall invoiced over $830,000 in charges to USC users, not including income derived from charges made to non-USC researchers in industry and academia. Over $2M in charges have been invoiced since 2017 by two core facilities alone housed at

Michelson Hall since its opening five years ago — the John O'Brien Nanofabrication Laboratory and the USC Center of Excellence in Nano Imaging.

Driving innovation: academia-private sector partnerships: Partnerships between academia and industry can drive innovation and have benefits for both parties. By enabling substantive and meaningful partnerships between academia and the private sector, Michelson Center core facilities build capacity for instrument development in academic settings and create new products with wide scientific and commercial use. Such mutually beneficial relationships produce ground-breaking research, drive innovation to solve complex problems and, importantly, propel economic growth.

Located at Michelson Hall since the core's opening in 2018, the Agilent Center of Excellence in Biomolecular Characterization is an example of such a collaboration and serves as the cornerstone of a larger partnership between USC and Agilent to accelerate fundamental discoveries in life sciences research. Michelson Center's Dynamic Imaging Science Center core facility is another example of academia/industry partnership between USC's Viterbi School of Engineering and Siemens "healthineers."

Educational outreach: Michelson Hall cores provide strong support to USC's educational mission, providing high-quality educational opportunities for both graduate and undergraduate students. Educational outreach programs enable cross-disciplinary research-intensive learning environments, integrate research and research

training, and empower the next generation of scientists. Student access to cutting-edge facilities creates a more skilled workforce, increasing the talent pool and strengthening the pipeline from research student to employee. Core facilities at the Michelson Center maintain close interaction with USC graduate students, such as lab tours of the Dynamic Imaging Center to Master's students from USC's Speech and Language Pathology program. Educational activities made available to local high school students are also an important component of core facility educational outreach, enabling cross-disciplinary research-intensive learning environments.

USC Support of Core Facilities

The USC Research Initiatives & Infrastructure (RII) office has recently launched a $2.15M Instrumentation Award program, providing financial assistance for the acquisition of major state-of-the-art equipment to core facilities, thereby improving access to, and increased use of modern research and research training instrumentation by a diverse workforce of scientists, engineers, and graduate and undergraduate students. The award program fosters the development of the next generation of instrumentation, resulting in new instruments that are more widely used, and/or open up new areas of research and research training.

The RII award program also facilitates the acquisition of new instrumentation by encouraging and supporting the submission of instrumentation grant proposals to federal agencies, such as the National Institutes of Health S10 program and the National Science Foundation Major Research Instrumentation (MRI) program, as well as to other federal and state agencies and private foundations. Dr. Jayakanth Ravichandran, Associate Professor of Chemical Engineering and Materials Science and co-director of the Center for Excellence in Nano Imagining facility was a recent awardee of the program, receiving $650,000 to be used towards the purchase of a Kratos AXIS Supra+ : XPS surface analysis instrument.
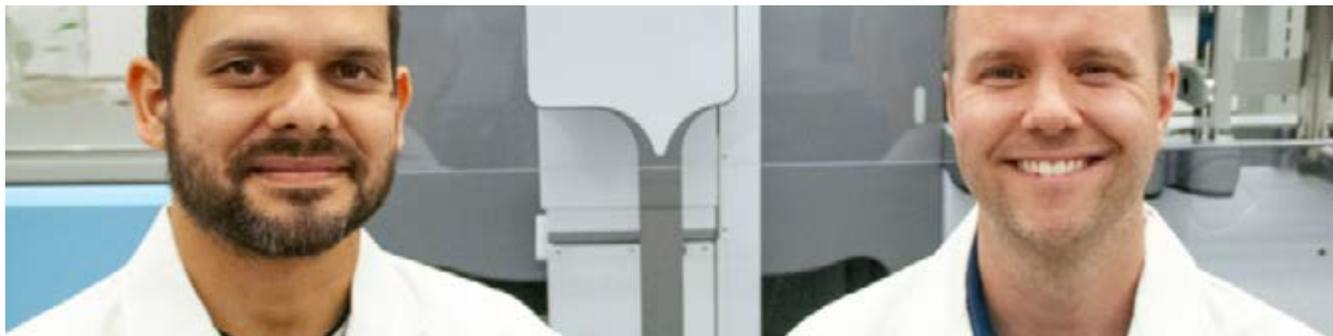
Agilent Center of Excellence in Biomolecular Characterization

Located at Michelson since 2018.
Visit the core on Youtube.

Core Directors:
Dr. Valery Fokin, Professor of Chemistry
Dr. Richard Roberts, Professor of Chemistry

The Agilent Center of Excellence in Biomolecular Characterization was established as a partnership between USC and Agilent technologies with the goal to converge researchers across science and engineering to work together on multidisciplinary approaches for the development of new drugs, diagnostics, and medical devices. The center features first-in-class research instruments ranging from advanced mass spectrometry, chromatography, robotic high-throughput reaction screening, genomics, and optical characterization.

The Agilent COE also supports the Agilent Fellows program. One Agilent Fellow is selected each year for a non-renewable two-year postdoctoral fellowship. The core will soon be announcing the selection of the 5th Agilent fellow to be hired in early 2023; the review of postdoc applications is currently underway. Drs. Alessandro Coradini and Matthew Sarnowski were selected for the 2022 Agilent Fellows Program. Dr. Coradini recently received a provisional patent for a technology that allows chromosome synthesis and replacement in eukaryotic cells (CReATiNG). His research *"Design and synthesis of a minimal yeast chromosome"* applies technology he developed to minimize a eukaryotic chromosome, aiming to map in unprecedented detail all essential genetic elements on a single eukaryotic chromosome and enable the

de novo synthesis of the first minimal eukaryotic chromosome. Determination of the universal minimal set of genes needed to sustain life is expected to contribute to our understanding of life, with applications in medicine and synthetic biology.

The Agilent COE also hosted several visits and seminars at the Michelson Center in 2022, including the *Los Angeles Metropolitan Mass Spectrometry (LAMMS) Discussion Meeting (Mass Spectrometry Bio Applications*, October 2022). This discussion and networking meeting was organized in collaboration with Agilent and with the support of MOBILion Systems. Speakers included Dr. Iain Campuzano, Senior Principal Scientist at Amgen, speaking on *"Mass Spectrometry in Biopharma: From Small Molecules to Multi-Specific Antibodies and Beyond"* and Professor Joseph Loo, Professor, Department of Biological Chemistry, David Geffen School of Medicine, and Department of Chemistry & Biochemistry at the University of California, Los Angeles (UCLA), speaking on *"Mass Spectrometry Au Naturel: A Tool for Structural Biology."*

John O'Brien Nanofabrication Laboratory

Located at Michelson since 2021. Opened for operations in 2022.
Visit the core on Vimeo.

Core Directors:
Dr. Rehan Kapadia, Director, Associate Professor of Electrical and Computer Engineering
Dr. Shiva Bhaskaran, Associate Director, Nanofabrication Laboratory

The vision behind the design of the John O'Brien Nanofabrication Laboratory is to move beyond conventional silicon processing. The breadth of instrumentation enables the core to support diverse research activities, from fundamental quantum transport studies to nano/bioconvergent research. Enabling interdisciplinary research with next generation nanomaterials is a strong focus of the facility. This emphasis is unique among nanofabrication facilities and enables researchers to translate novel materials into transformational devices for a wide range of applications.

Dr. Andrea Armani served as director of the core since its move to Michelson Hall in 2021 (and previously as the director of the USC Nanofab facilities, located in Viterbi, since 2017). As of January 2023, Dr. Rehan Kapadia will serve as the new core director. Dr. Kapadia's research interests lie at the intersection of material science and electrical engineering. Specifically, the next-generation electronic and photonic devices, as well as the material growth techniques necessary to fabricate these devices. Notably, he has developed new techniques for growing and integrating compound semiconductors with arbitrary substrates.

The core has seen extensive growth over the years, from 1.2 staff to a staff of five, with two additional hires in progress to support the growth of the user base. With over 150 +users, 40+ active research projects, and 35+ processing and characterization tools, the core's instrumentation and areas of research include: photonics, quantum, superconducting, bio devices, materials, sensors, MEMS, III V compound semiconductor, memristor, neuromorphic and hybrid computing, heterogenous integration, carbon nanotubes, neural interface, and battery technology.

Dynamic Imaging Science Center

Located at Michelson since 2021. Opened for operations in 2022.

DISC Management and Oversight Board:
Dr. Dani Byrd, Professor of Linguistics
Dr. Justin Haldar, Associate Professor of Electrical Engineering
Dr. Khalil Iskarous, Associate Professor of Linguistics
Dr. Shrikanth Narayanan, Professor and Niki and Max Nikias Chair in Engineering
Dr. Krishna Nayak, Professor, Viterbi ECE Systems

The mission of the Dynamic Imaging Science Center (DISC) is to better understand the science of human movement in health and disease through the development and use of non-invasive imaging. The center is supported by the National Science Foundation, the Viterbi School of Engineering, and Siemens "healthineers." It is home to an ultra-rare low field MRI machine, one of only three of its kind in the world, which is being used in research at Viterbi to develop "forever implants" to better understand when implants are failing, and to help manufacturers optimize new devices and doctors avoid unnecessary surgeries.

The facility is administered by a Management and Oversight Board, charged to review the activities of the Center and to consider and comment on plans for future activities. As well, a Technical Advisory Committee, comprised of USC faculty and a representative from Siemens, serves in an advisory capacity related to strategic planning and policy for technical operations.

The core facility hosted a number of events in 2022, including educational outreach activities, providing tours to groups from local high schools, helping to educate the next generation of scientists on the use of imaging in health and disease-related research. Educational outreach of the core also includes close interaction with Master's students from USC's Speech and Language Pathology program.

Current research at DISC is in the areas of Cardiac Perfusion, Fetal Imaging, Lung, Metallic Implants, and Vocal Production of Language.
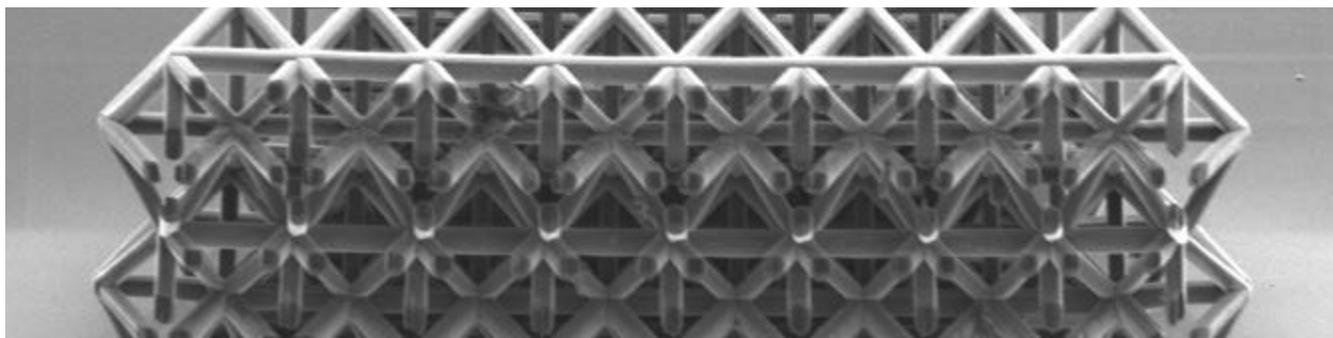
Center of Excellence in Nano Imaging

Located at Michelson since 2021.

Core Directors:
Dr. Jayakanth Ravichandran, Associate Professor of Chemical Engineering and Materials Science
Dr. James Boedicker, Associate Professor of Physics and Astronomy and Biological Sciences

The Center of Excellence in Nano Imaging (CNI) provides state-of-the-art imaging and characterization capabilities to the university and research communities. Its mission is to promote and advance the science and application of research tools for imaging, visualization, and analysis of nano- through meso-scale features and structures, both man-made and natural, including biological structures.

Dr. Ravichandran was awarded a $650,000 grant last December from USC's Research Initiatives and Infrastructure office for his proposal *"Chemical analysis from atomic to bulk length scales."* The award funded the purchase of a Kratos AXIS Supra+ : XPS surface analysis instrument, expanding the services and state-of-the-art instrumentation available to core facility users. The Viterbi School of Engineering and Dornsife College also provided funds towards the purchase of the equipment, to be housed at CNI.

Additional new equipment at the core includes: a MicroED system for the Krios cryoTEM, funded by Amgen, and a Thermo Fisher Talos F200C room temperature TEM with a Gatan K3 direct electron camera, donated by Amgen. As well, installation is in progress of a Thermo Fisher Spectra 200 Scanning transmission electron microscope with 4D STEM, EELS, Dual EDS detectors and STEM Tomography software.

Core services include new user training, scientific support and collaboration, as well services to industry and other external entities, in microscopy, X-Ray and sample preparation. The facility continues to grow is capabilities to keep USC at the forefront of research areas spanning physical to biological sciences.

Center for Peptide and Protein Engineering

Located at Michelson since 2018.

Core Directors:
Dr. Richard Roberts, Professor of Chemistry, Chemical Engineering and Materials Science and Biomedical Engineering.
Dr. Terry Takahashi, Assistant Research Professor of Chemistry

Peptides and proteins are the cornerstones of existing technology in biological recognition, diagnostics and human therapies. The CPPE is advancing the field of protein and peptide engineering, similar to the gains seen when the assembly line was introduced, by providing high-throughput generation of new reagents that recognize proteins and biological molecules.

Specialized services made available by the CPPE include a la carte and comprehensive peptide synthesis and purification. End-to-end production of peptides includes assistance in peptide design, Fmoc-based solid-phase peptide synthesis, peptide purification, conformation of peptide identity by mass spectrometry, and assessment of purity by reverse-phase HPLC. Peptide synthesis is performed on a Biotage Syro I Multi-Peptide Synthesizer (supported by a grant from the USC Core Instrumentation Fund) at 50 μmol scale up to 20 amino acids. Peptide purification is performed via Biotage reverse-phase flash chromatography and/or reverse-phase HPLC chromatography using Agilent HPLC and LCMS equipment.

The facility also provides *a la carte* and comprehensive protein expression and purification services. End-to-end production of proteins includes analysis of protein expression options, assistance in protein/construct design, subcloning, transformation of plasmids into E. Coli expression strains, expression of the protein, cell lysis, protein purification, and SDS-PAGE analysis. Protein purification options include Immobilized metal affinity chromatography (IMAC; typically Ni-NTA/His Tag), Ion exchange (both anion and cation), and size exclusion chromatography and are performed using a GE Äkta system.

The CPPE core provides USC researchers with new tools for biology, engineering, and biomedicine, placing the Michelson Center for Convergent Bioscience at the forefront of these new opportunities.

The need for more sophisticated lab equipment is rising even as research grants are shrinking... Centralized laboratories offer an alternative for researchers with a predilection for the latest technology.

Gould, J. Core facilities: Shared support. Nature 519, 495–496 (2015). https://doi.org/10.1038/nj7544-495a

Scientific & Social Impact

Current scientific challenges cut across the traditional boundaries of disciplines and the lifecycle of innovation — from research, to knowledge development, to its clinical application. Public understanding and engagement with science, and citizen participation — including through the popularization of science — are essential to equip citizens to make informed personal and professional choices. The scientific and social impact of the groundbreaking research being conducted at the Michelson Center is presented here — research that addresses societal needs and global health challenges.

Michelson Center

New to Michelson

We welcome new "Michelson Faculty" member Dr. Peter Foster, who joined USC in January 2023 as Assistant Professor of Physics and Astronomy. His research, in collaboration with Michelson's Bridge Institute, focuses on the interface between active matter physics and cell biology, with a focus on the cellular cytoskeleton. Dr. Foster holds a Bachelor of Science in Physics and Mathematics from the University of Iowa and a Ph.D. in Applied Physics from Harvard University.

Serving as the new director of the John O'Brien Nanofabrication Laboratory will be Dr. Rehan Kapadia. He received his Bachelor's degree in Electrical Engineering from the University of Texas at Austin, and his Ph.D. in Electrical Engineering from the

University of California, Berkeley. During his time at Berkeley, he was a *National Science Foundation Graduate Research Fellow*, and also won the *David J. Sakrison Memorial Prize* for outstanding research. Dr. Kapadia has also won an *AFOSR Young Investigator Award* (2016), an *ONR Young Investigator Award* (2021), and the *AVS Peter Mark Memorial Award* (2020). His research interests lie at the intersection of material science and electrical engineering, specifically, the next-generation electronic and photonic devices, as well as the material growth techniques necessary to fabricate these devices. Notably, he has developed new techniques for growing and integrating compound semiconductors with arbitrary substrates.

Peter Foster Joins Michelson Center

New Core Director Rehan Kapadia

Events

Celebrating of the 5-year anniversary of the Michelson Center for Convergent Bioscience, Dr. Gary Michelson toured Michelson Hall last November. He was accompanied by Geoffrey Baum, Executive Director of the Michelson Philanthropies & Michelson Center for Public Policy and Meyer Luskin, Chairman and CEO Scope Industries, and met with Dr. Ishwar K. Puri, USC Senior Vice President of Research & Innovation, and with Drs. Scott Fraser, Moh El-Naggar, Kate White, Krishna Nayak, Jayakanth Ravichandran and other Michelson Faculty to discuss ongoing and future research at the Center.

In the News

Recent research led by Dr. Vsevolod Katritch, in collaboration with scientists from Washington University in St. Louis and Stanford University, demonstrated that by chemically linking fentanyl to the sodium pockets that exist within nerve cell receptors, they could block the drug's harmful side effects and still reduce pain. While further studies are needed, the research, featured in a number of news outlets and published last November in the journal Nature, gives hope to address the current crisis of addiction and overdose deaths in the US, with over 68,000 opioid-related deaths in 2020 alone, according to the Centers for Disease Control.

Faculty

Andrea Armani

Events
Invited 2022 Keynote Speaker at the Kentucky Academy of Science (KAS) on *Traveling Through Disciplines*. Dr. Armani spoke of her academic career and the critical decision points which have led her to her current research, focused on the development of new diagnostic systems, synthesizing new materials and building new optical devices in order to better understand disease progression.

James Boedicker

Training the Next Generation
Provided mentoring to Ria Mandal, who presented her research project *"Array Expansion and Contraction with CRISPR"* as one of 51 members of the BUGS Jr Class at the 2022 BUGS Finale Symposium. Students presented on a broad range of topics, including biomedical imaging to natural products, genetic engineering of fungi, traumatic brain injury, neuropsychiatric

complications from COVID-19, among many others.

Dani Byrd

Events
Hosted Dr. Christina Hagedorn, City University of New York, at the Fall 2022 Otolaryngology and Hearing & Communication Neuroscience Seminar Series.

Led a PhonLunch: CV & Website workshop; PhonLunches are weekly gathering of individuals interested in the phonetic and phonological areas of Linguistics. The research-related focus of the meetings changes from week to week and also include topics dedicated to professional development.

Eun Ji Chung

Events
Was an invited speaker at University of Connecticut, Biomedical Engineering Department Seminar Series,

Gary K. Michelson Returns to Michelson Center

Redesigning Fentanyl

presented *Emerging Therapeutic Approaches in AKI: New Targets and How to Reach Them* at the American Society of Nephrology's ASN Kidney Week event.

Spoke at University of Massachusetts, Amherst, Department of Biomedical Engineering Seminar Series, presented *MicroRNA Delivery for Cardiovascular Disease*.

Spoke at Tufts University, Department of Biodemdical Engineering Seminar Series, presented *Biologically Inspired Nanoparticles for Chronic and Genetic Diseases*.

Stacey Finley

Honors & Awards
Presented with the *Cancer Warrior Award* from the University of Colorado Comprehensive Cancer Center in recognition of "outstanding contribution and achievement in cancer research and devotion to training future cancer warriors."

Events
Invited speaker at the University of Colorado, Comprehensive Cancer Center, Aurora, CO, the Vanderbilt University, Graduate Program in Cancer Biology and the City College of the City University of New York, Department of Chemical Engineering seminar seminar to present "*Computational modeling of the tumor ecosystem.*"

Organized the *Conference Committee for Systems Approaches to Cancer Biology* 2022 meeting, Woods Hole, MA on October 19-22, 2022. The ACSB is an organization that organically evolved from the Integrative Cancer Biology Program, an NIH National Cancer Institute-led initiative to promote systems biology approaches to cancer research.

In the News
Interviewed in Science & Technology News on the potential impact of the media to raise awareness on career opportunities in STEM among Black youth who might not otherwise consider a career in science — "*Will Black Panther: Wakanda Forever help make science more diverse?*"

Training the Next Generation
Authored the *"Data Science"* book, as a part of the JEP-sponsored program bringing free books to LAUSD students. To further empower and inspire, hundreds of the full book series were given to students in local schools near USC. Books were written in both English and Spanish allowing students to engage in the stories without a language barrier. Room to Read also created a series of videos to bring the books and scientists to life. Authors made lesson plans to accompany the books. The lessons help teachers bring the concepts from the books into the classroom.

Valery Fokin

Events
Led a Click Chemistry training seminar "*What I Talk About When I Talk About*" on proper use of vocabulary on a variety of research topics.

Scott Fraser

Honors & Awards
Awarded the *2022 Outstanding Researcher Award* from The Southern California Biomedical Council (SoCalBio).

Events
Presented at the UCLA David Geffen School of Medicine Cardiovascular Seminar *Eavesdropping on the Cellular and Molecular Interactions that Build Embryos*.

Khalil Iskarous

Events
Spoke at the Ludwig-Maximilians Universitat Institute of Phonetics and Speech Processing Methods and Approaches of Modern Phonetic Research on *Phonological Invariance through Phonetic Variation: The Dynamics of American English Intonation*.

Vsevolod Katritch

Honors & Awards
Awarded R01GM147537 (Katritch, PI) from NIGMS for *"Efficient synthon-based modular screening of Giga-to-Terra-scale virtual libraries"*. The project will continue the development of next-generation virtual screening technologies and apply them to the discovery of novel leads and drug candidates for a variety of clinically relevant targets.

In the News
Reported in Pain News Networks — *Can Fentanyl Be Made Safer?* New collaborative work suggests that the drug can be redesigned in such a way that this frequent overdose killer is converted to a much more benign but still effective analgesic. Also reported in Drugs of Abuse Update: A Safer Opioid and 3 New Biomarkers.

Steve Kay

Honors & Awards
Designated a *2022 Clarivate Highly Cited Researcher*, one of 10 USC faculty recognized by Clarivate for the demonstrated significant and broad influence of the research as reflected in their publication of multiple highly cited papers over the last decade.

Events
Presented *"Targeting Circadian Clock Proteins in Cancer"* at the University of Texas McGovern Medical School, Institute of Molecular Medicine 2022 seminar.

Co-chaired a retreat series, sponsored by the Keck School of Medicine and the Viterbi School of Engineering, to foster new and enhanced research collaborations between the two schools. Retreats aim to "leverage respective expertise, capabilities, and resources to enable the discovery and implementation of new solutions for unmet medical needs" with a "cross-disciplinary emphasis that integrates early discovery/experimental approaches, translational and applied research, data, enabling technologies, and cores."

In the News
Reported, a breakthrough in the treatment of Glioblastoma — a new class of small molecule drugs, known as SHP656, that can target the circadian clock proteins and may prove effective for treating the recurrence and spread of the disease.

Carl Kesselman

Honors & Awards
Named Institute of Electrical and Electronics Engineers' *2023 IEEE Fellow*, a distinction reserved for select IEEE members whose extraordinary accomplishments in any of the IEEE fields of interest are deemed fitting for this prestigious grade elevation. Kesselman was recognized for "foundational contributions to technologies and applications of global distributed computing." His research focuses on the application of advanced distributed computing and collaboration techniques to Grand Challenge problems in health informatics, biomedical science and policy. The rank of IEEE Fellow is the highest grade of membership at the organization, and less than 0.1% of voting members are selected annually for the IEEE member grade elevation.

Steve Kay Designated a 2022 Clarivate Researcher

Carl Kesselman Named IEEE Fellow

Awarded distinguished *IEEE Technical Field Award* for contributions to the design, deployment, and application of practical Internet-scale global computing platforms.

Yasser Khan

Training the Next Generation
Hosted 80 students and teachers from BIPOC (Black, Indigenous, and People of Color) middle and high school students in the Los Angeles area, providing activities focused on human-centered sensing and signal processing. Events held and developed in collaboration with USC Viterbi's K-12 STEM Center and the IEEE Signal Processing Society, provide activities, materials, and programs designed to prepare K-12 students for careers in STEM and educate them about the intersection of engineering and medicine.

Peter Kuhn

Honors & Awards
Awarded, together with Dr. James Hicks, a grant from the Breast Cancer Research Foundation for the project *Fluid Biopsy in Breast Cancer Patients to Characterize Cell Free and Cellular Constituents*, presented at the Breast Cancer Research Foundation Symposium & Awards Luncheon.

Events
Nominated to be a faculty participant and speaker at the inaugural *Future Vision Forum*, which aims to foster ideas and programs that accelerate the rate of innovation and discovery that improve the lives of people with eye disease and visual disability.

Hosted scholars, special advisors on military health, members of the Convergence and Advisory Board councils and invited speakers at the *1st Annual Workshop: Convergent Science in Precision Oncology* CSI-Cancer Center event, funded by a Department of Defense grant, Developing a Convergent Science Virtual Cancer Center.

Presented on *"Augmented Intelligence and Augmented Learning to Improve Patient Outcomes"* at the Cancer Research UK Lung Cancer Conference.

In the News
Featured in Diagnostics World, *Multianalyte Liquid Biopsy Detects Early-Stage Breast Cancer*, research by Dr. Kuhn and collaborators on the use of a high-definition liquid biopsy platform to stratify normal donors from those with early- or late-stage breast cancer.

Gianluca Lazzi

Events
Spoke at The 2022 International Symposium on Antennas and Propagation (ISAP), Sydney, Australia.

In the News
Featured in IEEE's Antennas & Propagation Magazine, President's Message, *"Looking Back at a Successful Year, IEEE Antenna and Propagation Society."*

Charles McKenna

In the News
Featured in Drug discovery, research by the McKenna lab to repair inner ear cells that are damaged from aging and prolonged exposure to noise.

Featured in Dentistry breakthrough, collaborative work with UCLA's Dr. Ichiro Nishimura and Dr. Varghese John, resulting in the co-invention of a topical formulation capable of delivering drugs to the jawbone through the gingival layer.

Ellis Meng

Honors & Awards
Awarded a $11.7 Million grant from the National Institutes of Health to create the Center for Autonomic Nerve Recording and Stimulation Systems (CARSS). CARSS will be

1st Annual CSVCC Workshop

$11.7M NIH Grant Creates CARSS

part of the *Human Open Research Neural Engineering Technologies (HORNET)* initiative within the *Stimulating Peripheral Activity to Relieve Conditions (SPARC)* program. The center will be co-led by investigators Victor Pikov from Medipace and Raja Hitti from Med-Ally and aims to develop cutting-edge devices and modular open-source implantable device systems to monitor and stimulate autonomic nerves inside the body.

Shrikanth Narayanan

Honors & Awards
Received a research award from the USC + Amazon Center on Secure and Trusted Machine Learning for the project *"Federated learning for human-centered experience and perception modeling with bio-behavioral data."* The awards support fundamental research and development of new approaches to machine learning (ML) privacy, security, and trustworthiness.

Events
Presented *"Leveraging Open Data and Task Augmentation to Automated Behavioral Coding of Psychotherapy Conversations in Low-Resource Scenarios"* at EMNLP 2022, one of the world's top natural language processing conferences.

Spoke at IEEE Education Society presenting *Pathways to Academia: Industry vs. Academia - Finding the Right Fit*.

In the News
Participated in the 2022 Emmys primetime presentation of the Governor's Award, with the Geena Davis Institute on Gender in Media, showcasing their work in developing media intelligence technologies in support of diversity and inclusion.

Remo Rohs

Honors & Awards
Elected American Association for the Advancement of Science (AAAS) Fellow in recognition of his contributions to science and technology.

Presented at Tel Aviv University, Tel Aviv, Israel on October 24, 2022.

Events
Spoke at the Weizmann Institute of Science Workshop, Rehovot, Israel, on *"Protein-DNA Interactions: From Biophysics to Cell Biology."*

Presented at the Max Planck Institute of Molecular Genetics, Berlin, Germany on October 27, 2022.

Maryam Shanechi

Events
Facilitated at *The Fifth Annual Bioelectronic Medicine Summit Decoding of Neural Signals* and the use of data science and machine learning in neural systems as applied to bioelectronic medicine. The session focused on efforts to understand how the computations performed by neurons relate to perception, movement, memory and other complex behaviors, and the decoding algorithms developed to translate sensory and motor neural signals into volition, perception, mental and disease states.

Presented at Stanford University's Wu Tsai Neurosciences Institute on *Dynamical Modeling, Decoding, and Control of Multiscale Brain Networks: from Motor to Mood*.

In the News
Received the *One Mind 2022 Rising Star Award* for her unique engineering approach to mental health. One Mind is a leading mental health non-profit that seeks to improve lives through science and advocacy. Shanechi was one of only three awardees

Fifth Bioelectronic Medicine Summit

Maryam Shanechi Receives Award

selected nationwide and whose research aims to develop AI algorithms that enable personalized therapy.

Kate White

Events
Organized and managed the Pancreatic Beta Cell Consortium Quarterly meeting in November. Key participants in the Consortium from the USC Michelson Center include Drs. Carl Kesselman, Stacey Finley, Ray Stevens, Vadim Cherezov, Valery Fokin, Helen Berman, and Scott Fraser. The Consortium focuses on developing new approaches for multi-scale imaging of pancreatic beta cells and harnessing newly developed modeling strategies to integrate multi-scale data.

Presented a lecture at the Cell Bio 2022 Meeting (ASCB/EMBO Meeting) in Washington DC.

Cristina Zavaleta

Events
Invited to the White House Initiative Summit to promote higher education among Hispanic students and the importance of advancing educational equity.

Invited to speak at Cal State L.A. Minority Opportunities in Research (MORE) program about her research and academic journey. The program makes available to the American scientific research enterprise the creativity and intellectual talents of minority group members.

Participated in Department of State's International Visitor Leadership Program, hosting global Women Leaders in STEM, providing a tour of her lab and the opportunity to showcase her ongoing research.

Training the Next Generation
Presented her research to REACH students from underserved high schools in the greater Los Angeles community.

Facilitated at the California Forum for Diversity in Graduate Education, speaking on her research and academic journey to underrepresented minority students interested in pursuing higher education and applying to graduate school.
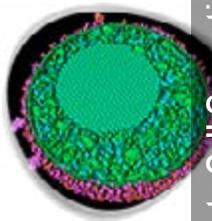
Pancreatic Beta Cell Consortium

Presented Research to REACH Students
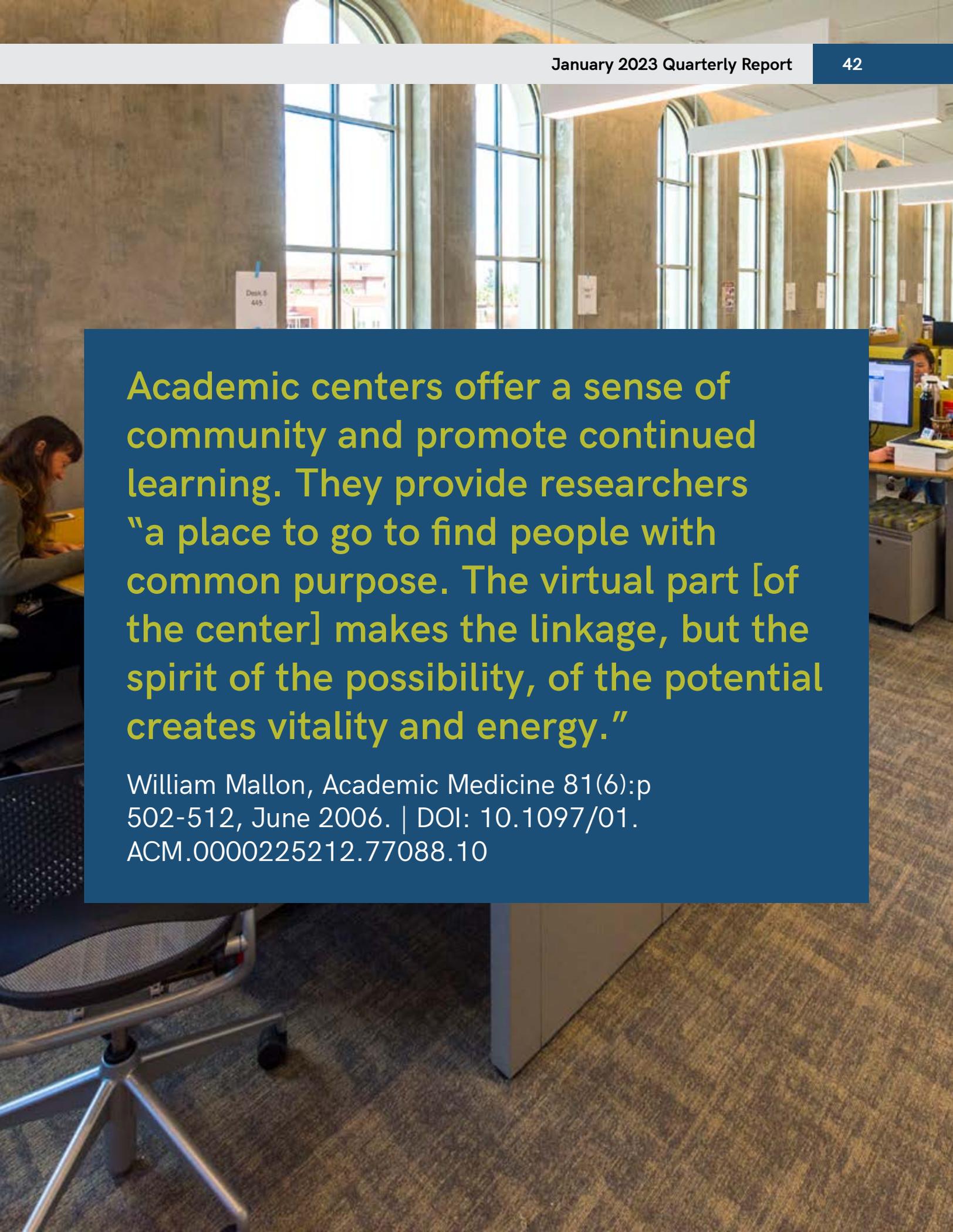
Academic centers offer a sense of community and promote continued learning. They provide researchers "a place to go to find people with common purpose. The virtual part [of the center] makes the linkage, but the spirit of the possibility, of the potential creates vitality and energy."

Manuscripts & Publications

It is through publication of journal articles and books that research is disseminated to the scientific community and the public at large, helping to advance new knowledge and facilitate its application. Listed below are highlights of Michelson faculty publications.

Vadim Cherezov

High ligand efficiency quinazoline compounds as novel A2A adenosine receptor antagonists. Eur J Med Chem. 2022 Nov 5;241:114620. https://pubmed.ncbi.nlm.nih.gov/35933788/

Eun Ji Chung

The oral delivery of kidney targeting nanotherapeutics for Polycystic Kidney Disease. bioRxiv pre-print: October 21, 2022. https://www.biorxiv.org/content/10.1101/2022.10.18.512444v1

There are Beneficial Effects of Bempedoic Acid Treatment in Polycystic Kidney Disease Cells and Mice. Frontiers in Molecular Biosciences. bioRxiv pre-print: October 8, 2022. https://www.biorxiv.org/content/10.1101/2022.08.09.503392v2

Stacey Finley

Systematic Bayesian posterior analysis guided by Kullback-Leibler divergence facilitates hypothesis formation. J Theor Biol. November 3, 2022. Online ahead of print. https://pubmed.ncbi.nlm.nih.gov/36335999/

Kinetic and data-driven modeling of pancreatic □-cell central carbon metabolism and insulin secretion. PLOS. Oct 17, 2022. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1010555

Rehan Kapadia

Monolithic III–V on Metal for Thermal Metasurfaces. ACS Publications. November 3, 2022. https://pubs.acs.org/doi/abs/10.1021/acsnano.2c06703

Vsevolod Katritch

Structure-based design of bitopic ligands for the µ-opioid receptor. Nature. Online ahead of print. November 30, 2022. https://pubmed.ncbi.nlm.nih.gov/36450356/

Germline selection of PTPN11 (HGNC:9644) variants make a major contribution to both Noonan syndrome's high birth rate and the transmission of sporadic cancer variants resulting in fetal abnormality. Wiley Online Library. November 9, 2022. https://doi.org/10.1002/humu.24493

Carl Kesselman

FaceBase: a community-driven hub for data-intensive research. Journal of Dental Research, 101(11), pp.1289-1298. October 2022. https://pubmed.ncbi.nlm.nih.gov/35912790/

Mercedeh Khajavikhan

Complex skin modes in non-Hermitian coupled laser arrays. Nature. November 28, 2022. https://www.nature.com/articles/s41377-022-01030-0

Peter Kuhn

Defining A Liquid Biopsy Profile of Circulating Tumor Cells and Oncosomes in Metastatic Colorectal Cancer for Clinical Utility. PubMed. October 6, 2022. https://pubmed.ncbi.nlm.nih.gov/36230811/

Characterization of BCMA Expression in Circulating Rare Single Cells of Patients with Plasma Cell Neoplasms. PubMed. November 3, 2022. https://pubmed.ncbi.nlm.nih.gov/36362214/

Jeremy Mason

A quantitative characterization of the spatial distribution of brain metastases from breast cancer and respective molecular subtypes. PubMed. October 16, 2022. https://pubmed.ncbi.nlm.nih.gov/36245013/

Characterization of BCMA Expression in Circulating Rare Single Cells of Patients with Plasma Cell Neoplasms. Int J Mol Sci. November 3, 2022. https://www.mdpi.com/1422-0067/23/21/13427

Defining A Liquid Biopsy Profile of Circulating Tumor Cells and Oncosomes in Metastatic Colorectal Cancers (Basel). MDPI. October 6, 2022. doi: 10.3390/cancers14194891

Charles McKenna

Toward more potent imidazopyridine inhibitors of Candida albicans Bdf1: Modeling the role of structural waters in selective ligand binding. PubMed. December 15, 2022. https://pubmed.ncbi.nlm.nih.gov/36190786/

Shrikanth Narayanan

Phone duration modeling for speaker age estimation in children. J Acoust Soc Am. November 15, 2022. https://asa.scitation.org/doi/abs/10.1121/10.0015198

Krishna Nayak

Contrast-optimal simultaneous multi-slice bSSFP cine cardiac imaging at 0.55 T. Wiley Online Library. October 5, 2022. https://onlinelibrary.wiley.com/doi/full/10.1002/mrm.29472

Lung parenchyma transverse relaxation rates at 0.55 T. Wiley Online Library. November 20, 2022. https://onlinelibrary.wiley.com/doi/full/10.1002/mrm.29541

Remo Rohs

Top-Down Crawl: a method for the ultra-rapid and motif-free alignment of sequences with associated binding metrics. Bioinformatics. November 15 2022. https://academic.oup.com/bioinformatics/article/38/22/5121/6731921

It is in the flanks: Conformational flexibility of transcription factor binding sites. Biophysical Journal. October 18, 2022. https://www.cell.com/biophysj/fulltext/S0006-3495(22)00770-6 18

Physicochemical models of protein-DNA binding with standard and modified bases. USA (in press). Proc. Natl. Acad. Sci. December 12, 2022. https://www.pnas.org/doi/10.1073/pnas.2205796120

Cristina Zavaleta

Hybrid Unmixing for longitudinal in vivo imaging of low signal to noise fluorescence. Nature Methods. Accepted October 2022. https://www.nature.com/articles/s41592-022-01751-5
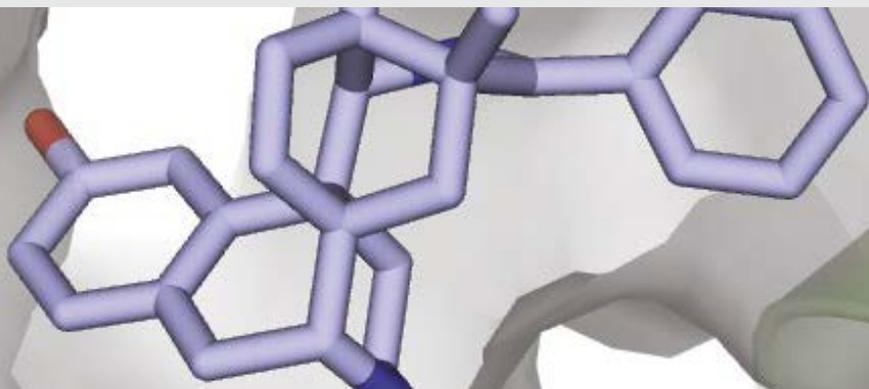
Featured Article

Structure-based design of bitopic ligands for the μ-opioid receptor

Vsevolod Katritch, Nature — November 30, 2022

Opioid receptors, members of the class A family of G-protein-coupled receptors (GPCRs), are key molecular targets in pain management. The challenge remains to engage specific subtypes of opioid receptor and to trigger a specific functional response that can produce in vivo analgesia without side effects such as tolerance, respiratory depression and addiction. Some approaches to mitigate these adverse reactions include the development of biased μOR agonists, peripherally restricted agonists, ligands targeting μOR splice variants, opioids with mixed actions at other subtypes and compounds that bind to μOR only in acidic environments. Most of these past approaches have targeted the orthosteric site, although positive allosteric modulators have recently been reported at μORs. Here we aimed to target a distinct Na+-binding allosteric site, which is highly conserved in a vast majority of family A GPCRs and is detected crystallographically in high-resolution inactive state structures and biochemically in at least diverse family A GPCRs.

Allosteric modulation of Gi and Go GPCR subtypes by sodium was first observed in the 1970s at the μOR. The presence of physiological NaCl concentrations was found to shift the receptor towards the inactive state, thus reducing agonist binding. As recent structural studies reveal, the pocket undergoes marked conformational changes upon receptor activation and is critical for the modulation of signalling in family A GPCRs. It has also been shown that the Na+-binding site residues serve as major 'efficacy switches' that can bias the GPCR signalling to either Gi protein or β-arrestin-2 pathways. The polar cavity with the Na+ site has also been characterized by molecular dynamics in inactive μOR, whereas the high-resolution structure of μOR bound to the agonist Bu72 (Protein Data Bank (PDB): 5C1M) revealed activation-related conformational changes in the cavity that prevent sodium binding but retain several water molecules.

Moreover, this structure revealed a water-filled polar channel linking the orthosteric pocket with the Na+ site, raising the possibility of developing ligands that interact directly with the Na+ site. The high conservation and key functional role of this allosteric pocket suggest that it is suitable as a target for allosteric modulators and bitopic ligands with unique properties for some family A GPCRs, including the μOR. To address this hypothesis, we used a structure-based approach and

designed a series of bitopic ligands on the fentanyl scaffold. These bitopic ligands target both the orthosteric pocket of the µOR and extend into the polar channel towards the Na^+ site. The binding pose of two fentanyl bitopic ligands was confirmed by structure determination using cryo-electron microscopy (cryo-EM). Functional characterization of these ligands revealed that the extension of a positively charged guanidine group into the Na^+ site resulted in a near complete loss of arrestin recruitment, and, surprisingly, substantial changes in the relative potency and efficacy for the six G_i, G_o and G_z subtypes. Finally, the lead bitopic ligand, C6 guano, showed antinociception with reduced side effects compared with classical µOR agonists such as morphine or the lethal agonist fentanyl. Owing to the high conservation of the Na^+-binding pocket, similar bitopic ligand designs may be suitable for effective control of functional selectivity in other class A GPCRs.

Bitopic µOR ligands based on the fentanyl scaffold

Fentanyl is a synthetic µOR-selective agonist that is 100-fold more potent as an analgesic than morphine. Unlike morphine, fentanyl also is highly efficacious for inducing arrestin translocation and is a full agonist for Gi-mediated signalling, whereas morphine is a partial agonist. Fentanyl also acts rapidly and is commonly used to manage postoperative and severe cancer-related pain, and is known to bind specifically to the orthosteric site of µOR—but in the USA, its use has led to around 60,000 deaths due to overdose.

Here we predicted the binding poses of fentanyl on the basis of the crystal structure of the active state µOR with the N-terminus truncated at residue M65, and computationally designed and synthesized a small library of fentanyl derivatives designed to extend through the polar channel below the orthosteric pocket to the Na^+-binding pocket. Although alternative models for fentanyl

binding have been suggested, the docking model used here has been validated by cryo-EM with the fentanyl analogue lofentanil, as well as by cryo-EM complexes with the designed bitopic ligand reported here. The models predicted the distance between the amide nitrogen of fentanyl and the carboxy group of the D114[2.50] residue of the allosteric Na^+-binding site to be approximately 13 Å (superscripts indicate Ballesteros–Weinstein numbering for GPCRs). In order to engage the allosteric Na^+-binding site, we replaced the classical fentanyl benzene ring with an aliphatic chain linker (C_n, where n = 3, 5, 6, 7, 9 or 11) connected to a positively charged 'warhead', an amine or guanidine (guano) group. The bitopic compound with a C7 linker (C7 guano) was predicted to be optimal for the formation of a direct salt bridge between the amino warhead and D114[2.50] carboxylic acid, and the C6 linker (C6 guano) was predicted to be optimal for the interaction for the guanidine warhead. The syntheses of amino and guano bitopic ligands31 on the fentanyl scaffold are presented in Supplementary Fig. 1.

Full article can be found in the addendum.
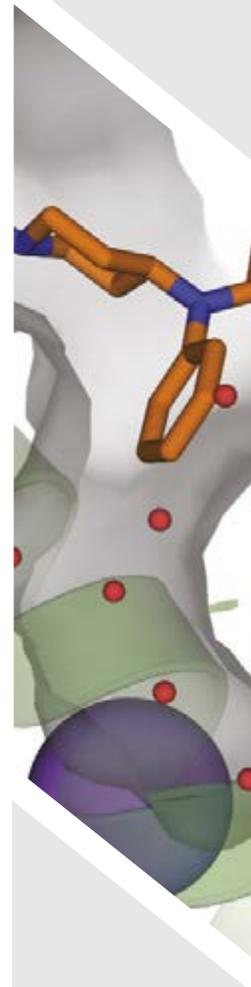
Highlighted Publications

Defining A Liquid Biopsy Profile of Circulating Tumor Cells and Oncosomes in Metastatic Colorectal Cancer for Clinical Utility

Peter Kuhn, Cancers — October 6, 2022

Metastatic colorectal cancer (mCRC) is characterized by its extensive disease heterogeneity, suggesting that individualized analysis could be vital to improving patient outcomes. As a minimally invasive approach, the liquid biopsy has the potential to longitudinally monitor heterogeneous analytes. Current platforms primarily utilize enrichment-based approaches for epithelial-derived circulating tumor cells (CTC), but this subtype is infrequent in the peripheral blood (PB) of mCRC patients, leading to the liquid biopsy's relative disuse in this cancer type. In this study, we evaluated 18 PB samples from 10 mCRC patients using the unbiased high-definition single-cell assay (HDSCA). We first employed a rare-event (Landscape) immunofluorescence (IF) protocol, which captured a heterogenous CTC and oncosome population, the likes of which was not observed across 50 normal donor (ND) samples. Subsequent analysis was conducted using a colorectal-targeted IF protocol to assess the frequency of CDX2-expressing CTCs and oncosomes. A multi-assay clustering analysis isolated morphologically distinct subtypes across the two IF stains, demonstrating the value of applying an unbiased single-cell approach to multiple assays in tandem. Rare-event enumerations at a single timepoint and the variation of these events over time correlated with progression-free survival. This study supports the clinical utility of an unbiased approach to interrogating the liquid biopsy in mCRC, representing the heterogeneity within the CTC classification and warranting the further molecular characterization of the rare-event analytes with clinical promise.

Physicochemical models of protein–DNA binding with standard and modified base pairs

Remo Rohs, PNAS — January 19, 2023

DNA-binding proteins play important roles in various cellular processes, but the mechanisms by which proteins recognize genomic target sites remain incompletely understood. Functional groups at the edges of the base pairs (bp) exposed in the DNA grooves represent physicochemical signatures. As these signatures enable proteins to form specific contacts between protein residues and bp, their study can provide mechanistic insights into protein–DNA binding. Existing experimental methods, such as X-ray crystallography, can reveal such mechanisms based on physicochemical interactions between proteins and their DNA target sites. However, the low throughput of structural

biology methods limits mechanistic insights for selection of many genomic sites. High-throughput binding assays enable prediction of potential target sites by determining relative binding affinities of a protein to massive numbers of DNA sequences. Many currently available computational methods are based on the sequence of standard Watson–Crick bp. They assume that the contribution of overall binding affinity is independent for each base pair, or alternatively include dinucleotides or short k-mers. These methods cannot directly expand to physicochemical contacts, and they are not suitable to apply to DNA modifications or non-Watson–Crick bp. These variations include DNA methylation, and synthetic or mismatched bp. The proposed method, DeepRec, can predict relative binding affinities as function of physicochemical signatures and the effect of DNA methylation or other chemical modifications on binding. Sequence-based modeling methods are in comparison a coarse-grain description and cannot achieve such insights. Our chemistry-based modeling framework provides a path towards understanding genome function at a mechanistic level.

Faculty At a Glance

Source of data, tables and graphs: Digital Science's Dimensions platform.

Andrea Armani

Publications	Datasets	Grants	Patents
212	1	15	14
Citations 3,637		Funding amount USD 7.6 M	Clinical Trials 0
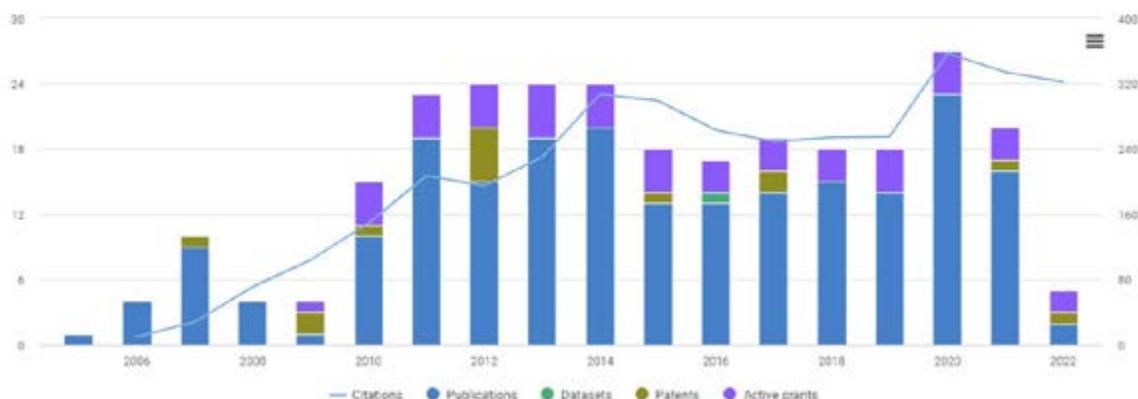
Publications: Collaborations
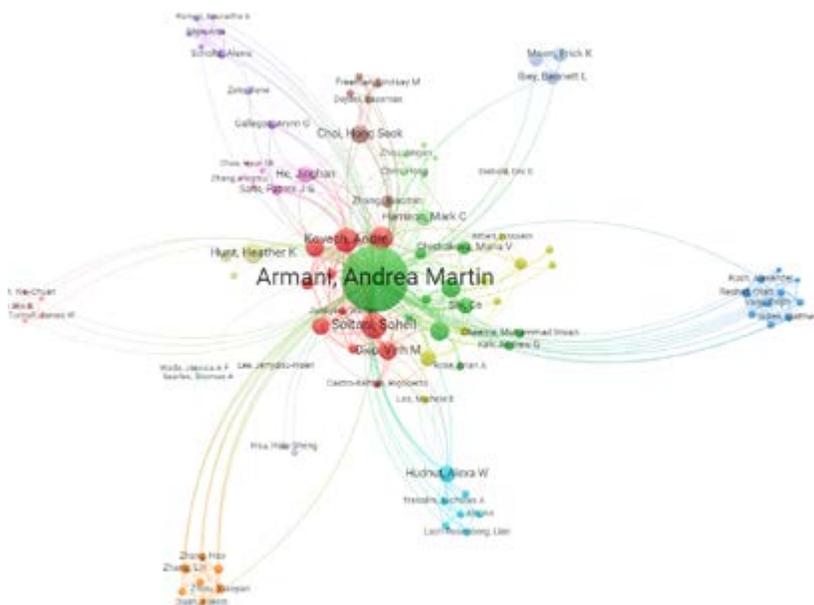
James Boedicker

Publications	Datasets	Grants	Patents
63	13	6	11
Citations		Funding amount	Clinical Trials
1,971		USD 8.8 M	0

Publications: Collaborations

Dani Byrd

Publications	Datasets	Grants	Patents
117	27	4	0
Citations		Funding amount	
2,773		USD 8.9 M	Clinical Trials
			0

Publications: Collaborations

Vadim Cherezov

Publications	Datasets	Grants	Patents
203	7	15	1
Citations		Funding amount	
28,690		USD 71.7 M	Clinical Trials
			0

Publications: Collaborations

Eun Ji Chung

Publications	Datasets	Grants	Patents
55	1	5	0
Citations 1,514		Funding amount USD 3.6 M	Clinical Trials 0

Publications: Collaborations

Mohamed El-Naggar

Publications	Datasets	Grants	Patents
88	7	2	3
Citations		Funding amount	
3,333		USD 2.0 M	Clinical Trials
			0

Citations — Publications ● Datasets ● Patents ● Active grants
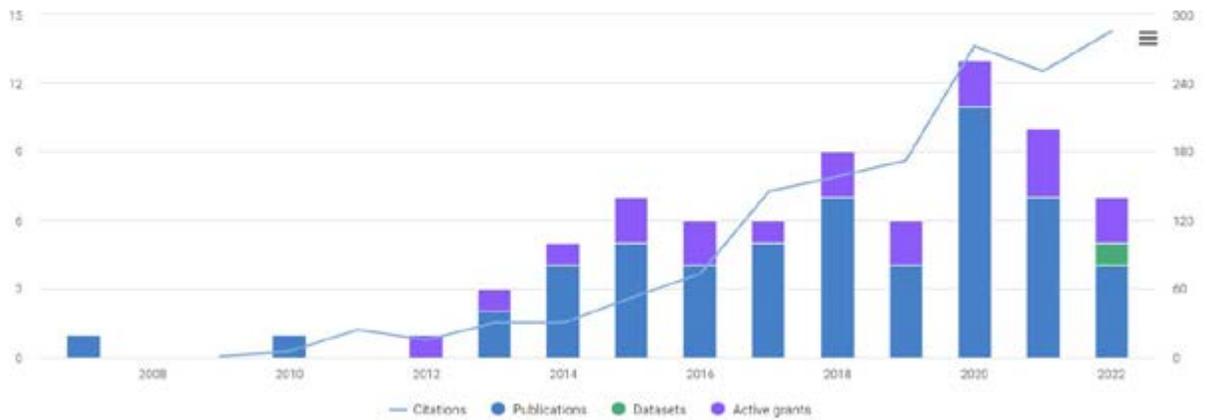
Publications: Collaborations
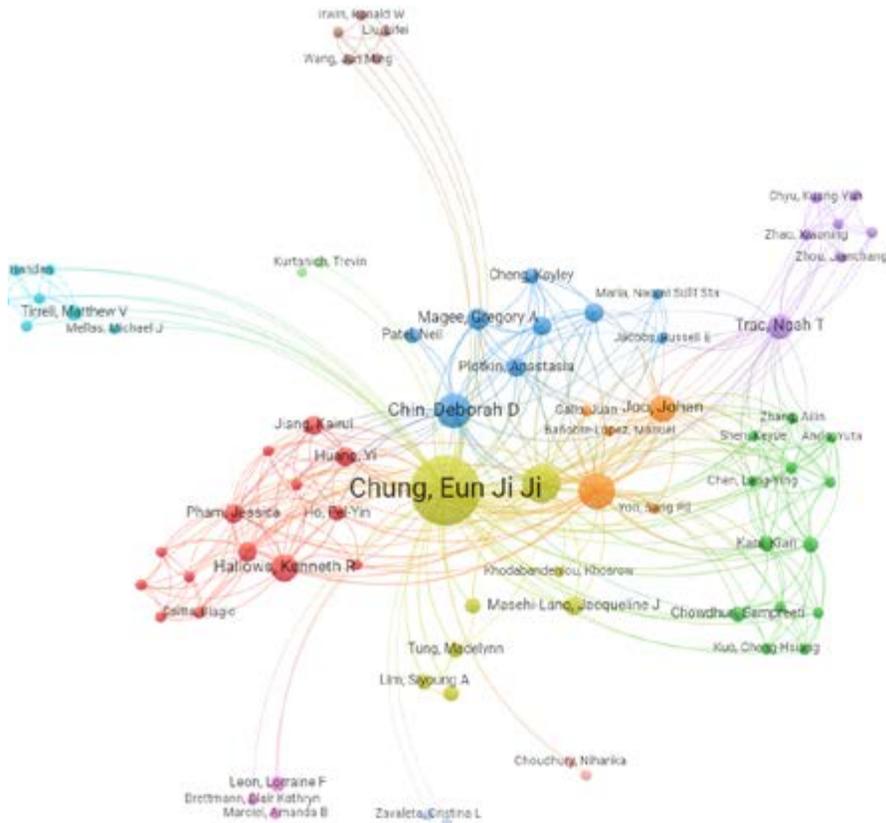
Stacey Finley

Publications	Datasets	Grants	Patents
86	19	4	0
Citations		**Funding amount**	
1,387		USD 4.0 M	**Clinical Trials**
			0

— Citations ● Publications ● Datasets ● Active grants

Publications: Collaborations

Valery Fokin

Publications	Datasets	Grants	Patents
249	**65**	**7**	**14**
Citations		Funding amount	
38,739		USD 33.6 M	Clinical Trials
			0

— Citations   ● Publications   ● Datasets   ● Patents   ● Active grants

Publications: Collaborations

Scott Fraser

Publications	Datasets	Grants	Patents
405	9	36	288
Citations		Funding amount	Clinical Trials
29,322		USD 105.2 M	0

Publications: Collaborations

Cornelius Gati

Publications	Datasets	Grants	Patents
70	1	1	1

Citations	Funding amount		Clinical Trials
4,813	0		0
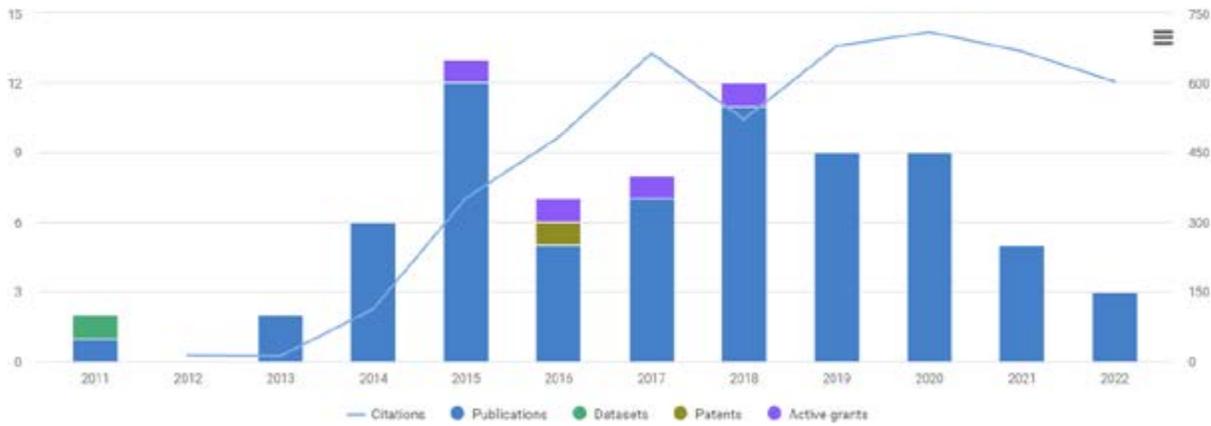
Publications: Collaborations
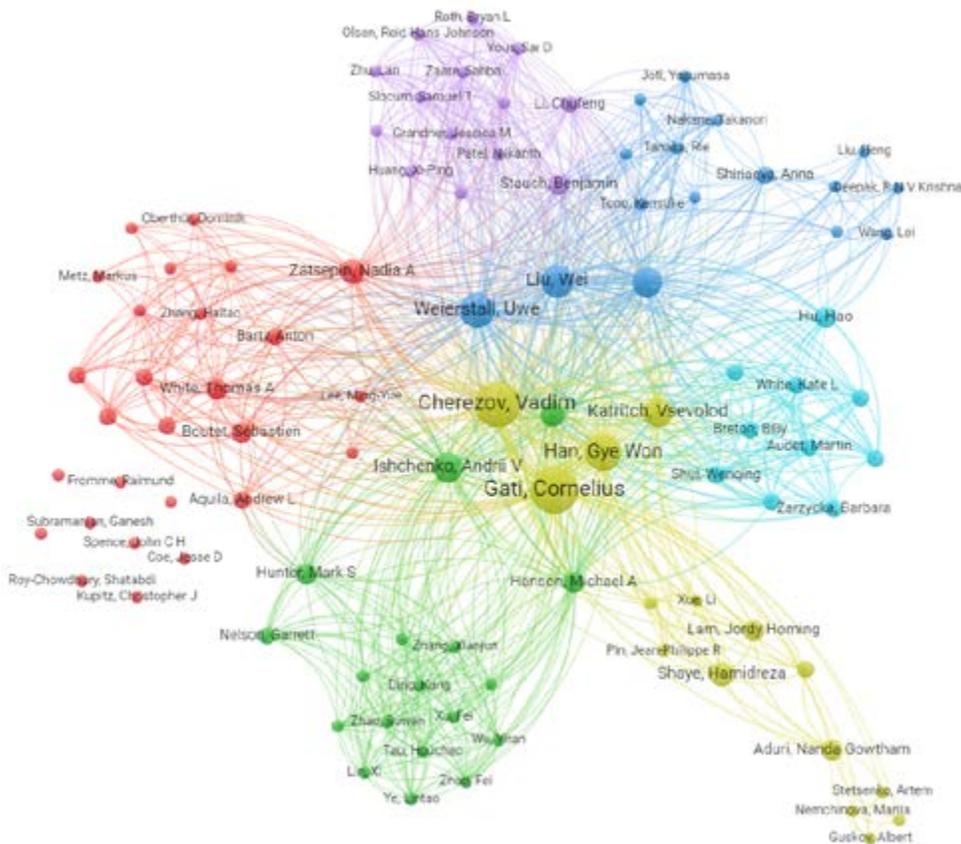
Justin Haldar

Publications	Datasets	Grants	Patents
131	2	3	7
Citations		Funding amount	
3,729		USD 3.3 M	Clinical Trials
			0

Publications: Collaborations

James Hicks

Publications	Datasets	Grants	Patents
218	8	14	4
Citations		Funding amount	
24,383		USD 4.6 M	Clinical Trials
			0
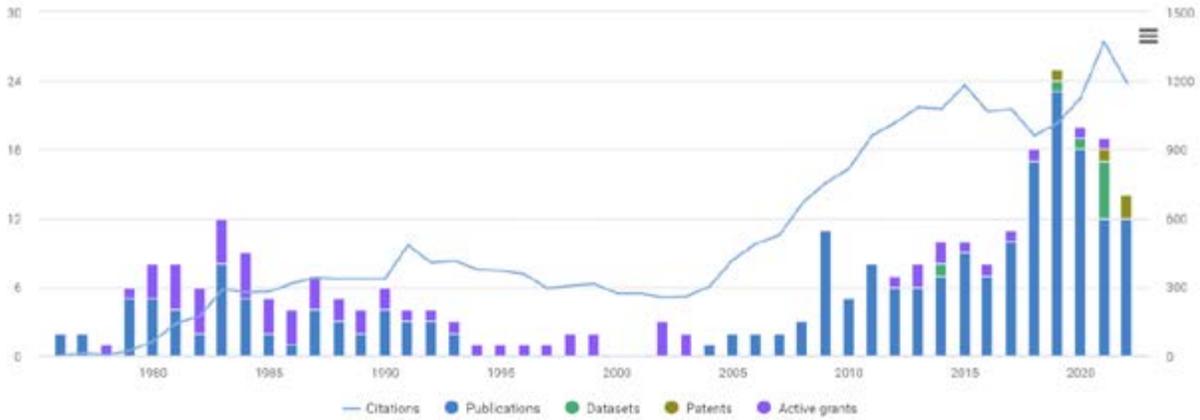
Publications: Collaborations
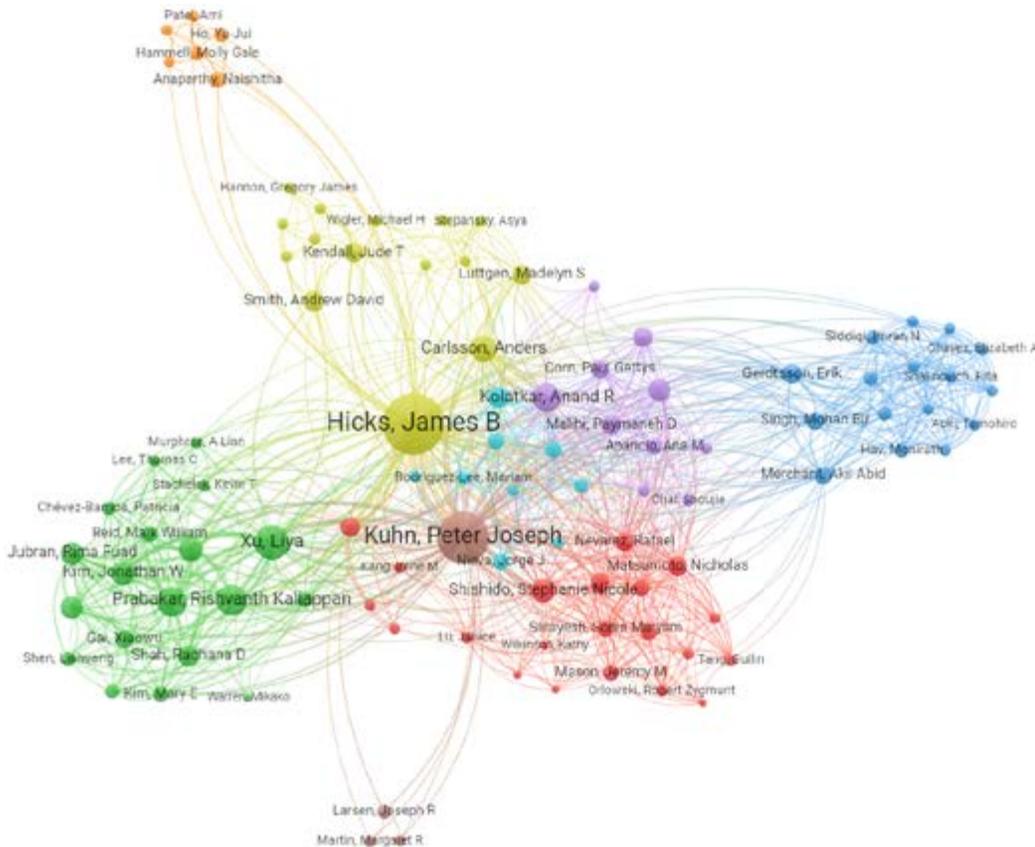
Khalil Iskarous

Publications	Datasets	Grants	Patents
65	0	5	0
Citations		Funding amount	
580		USD 3.5 M	Clinical Trials
			0

Publications: Collaborations

Rehan Kapadia

Publications	Datasets	Grants	Patents
105	2	6	7
Citations		Funding amount	
3,969		USD 2.6 M	Clinical Trials
			0
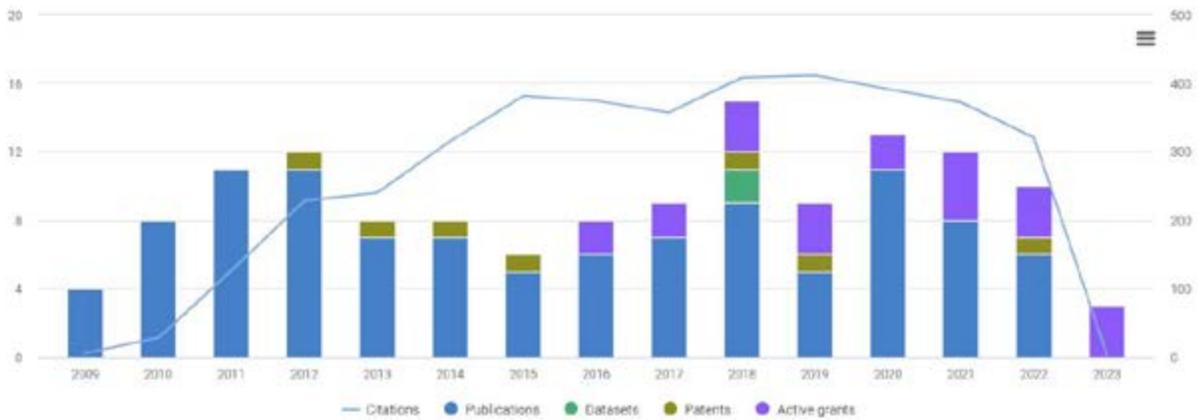
Publications: Collaborations
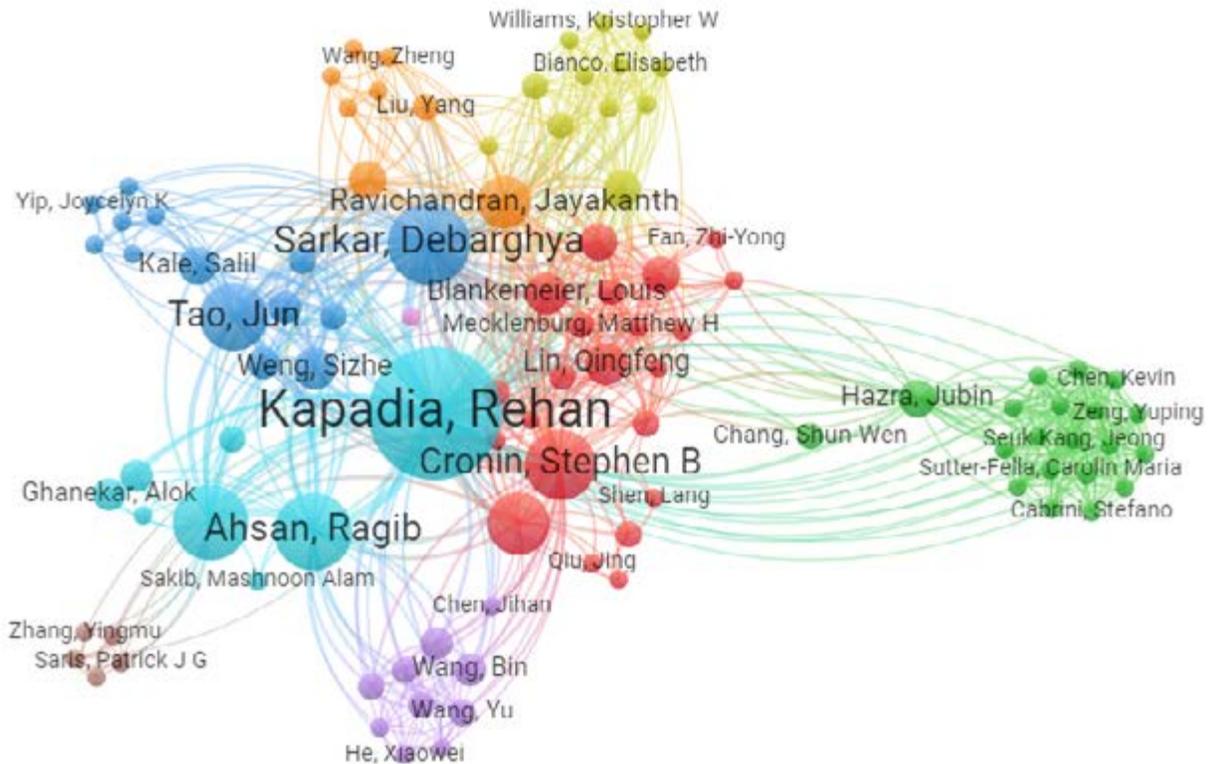
Vsevolod Katritch

Publications 147	Datasets 24	Grants 12	Patents 6
Citations 20,430		Funding amount USD 19.3 M	Clinical Trials 0
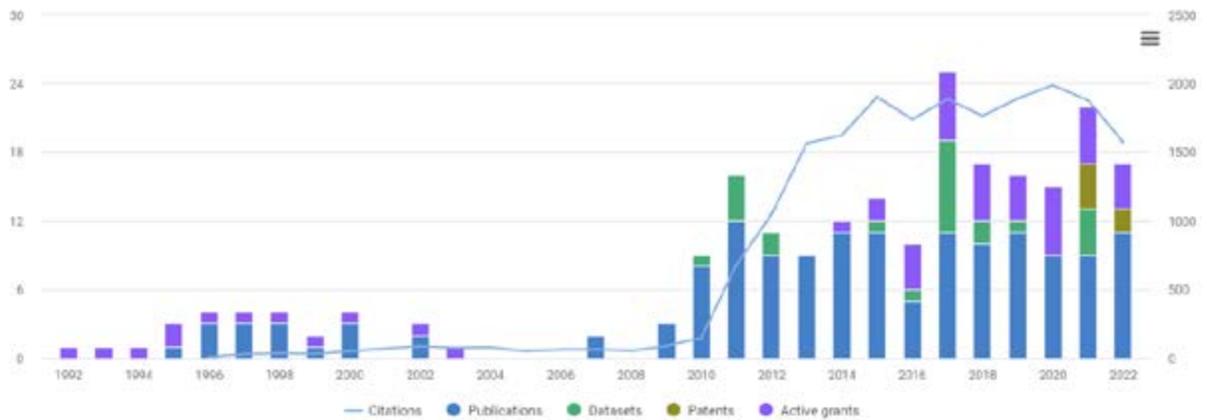
Publications: Collaborations
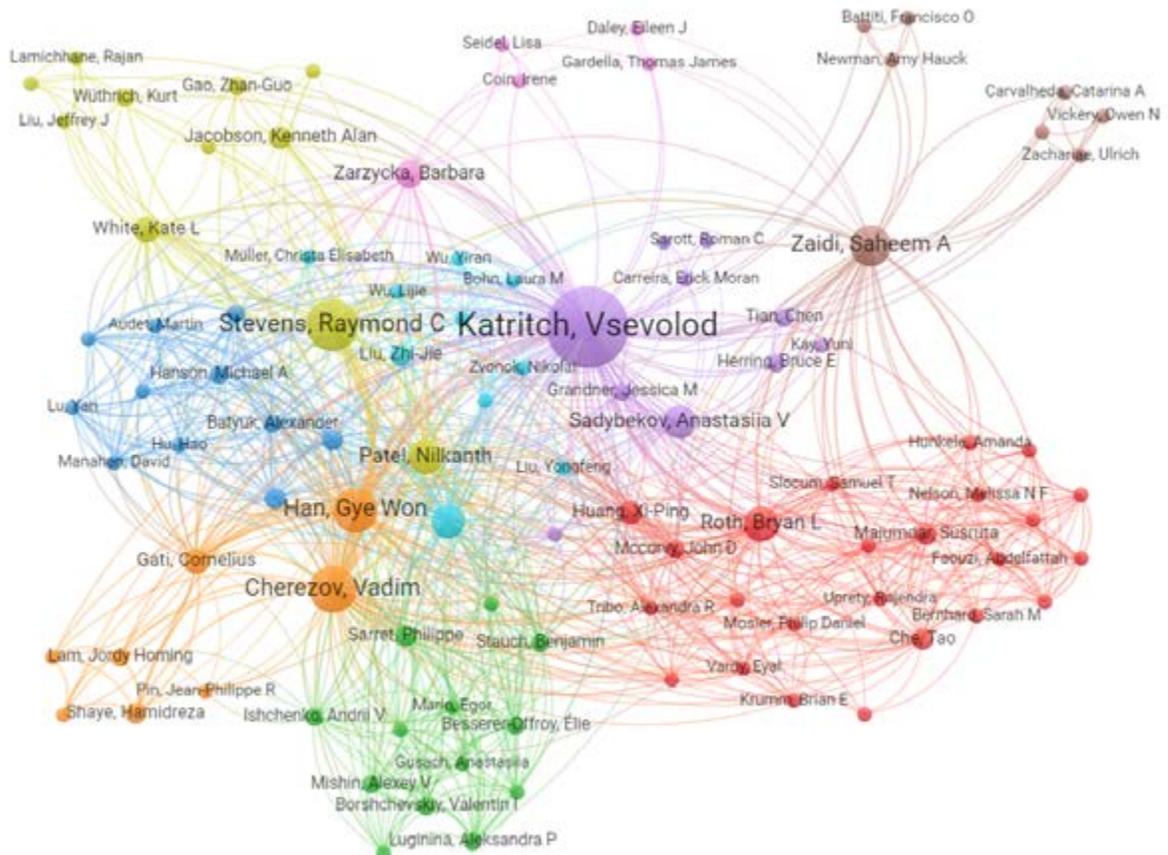
Steve Kay

Publications	Datasets	Grants	Patents
285	39	23	7
Citations		Funding amount	Clinical Trials
43,389		USD 34.6 M	0
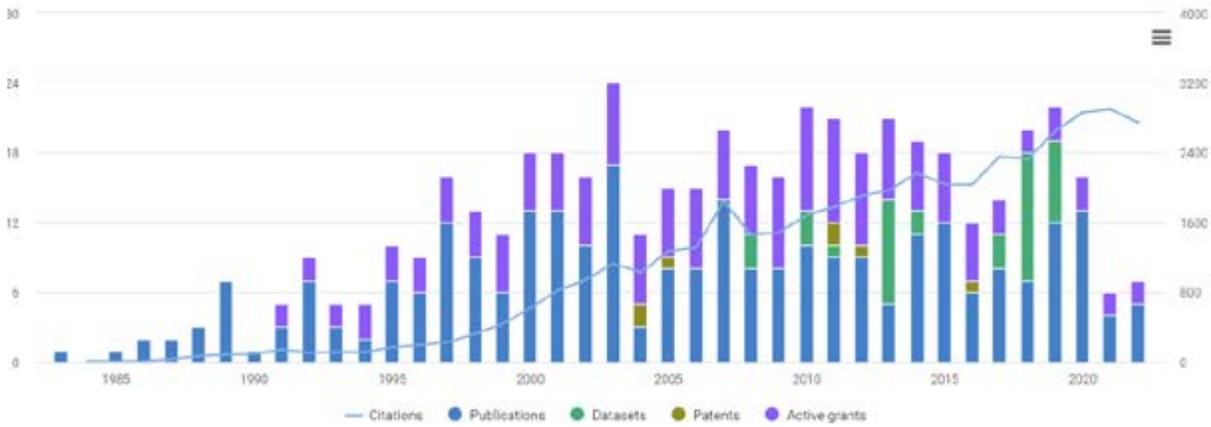
Publications: Collaborations
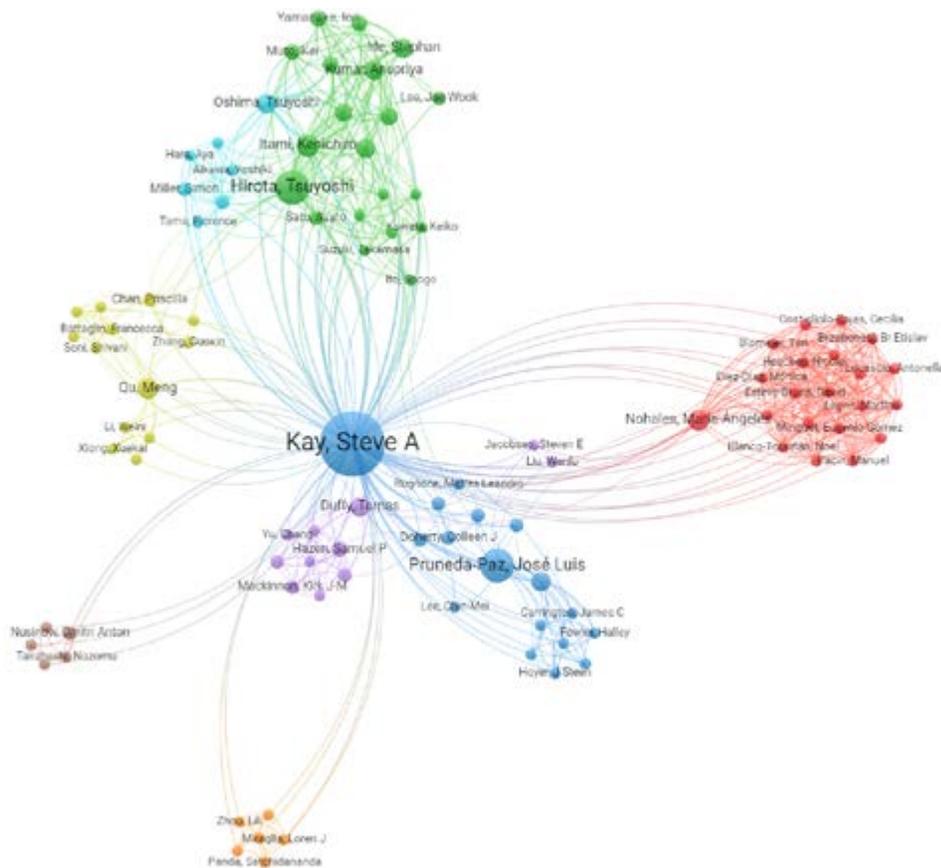
Carl Kesselman

Publications	Datasets	Grants	Patents
203	14	35	0
Citations		Funding amount	
21,822		USD 145.0 M	Clinical Trials
			0

Publications: Collaborations

Mercedeh Khajavikhan

Publications	Datasets	Grants	Patents
232	0	13	4
Citations		Funding amount	
7,220		USD 8.6 M	Clinical Trials
			0

Publications: Collaborations

Peter Kuhn

Publications	Datasets	Grants	Patents
316	20	15	28
Citations 17,726		Funding amount USD 321.4 M	Clinical Trials 2

Publications: Collaborations
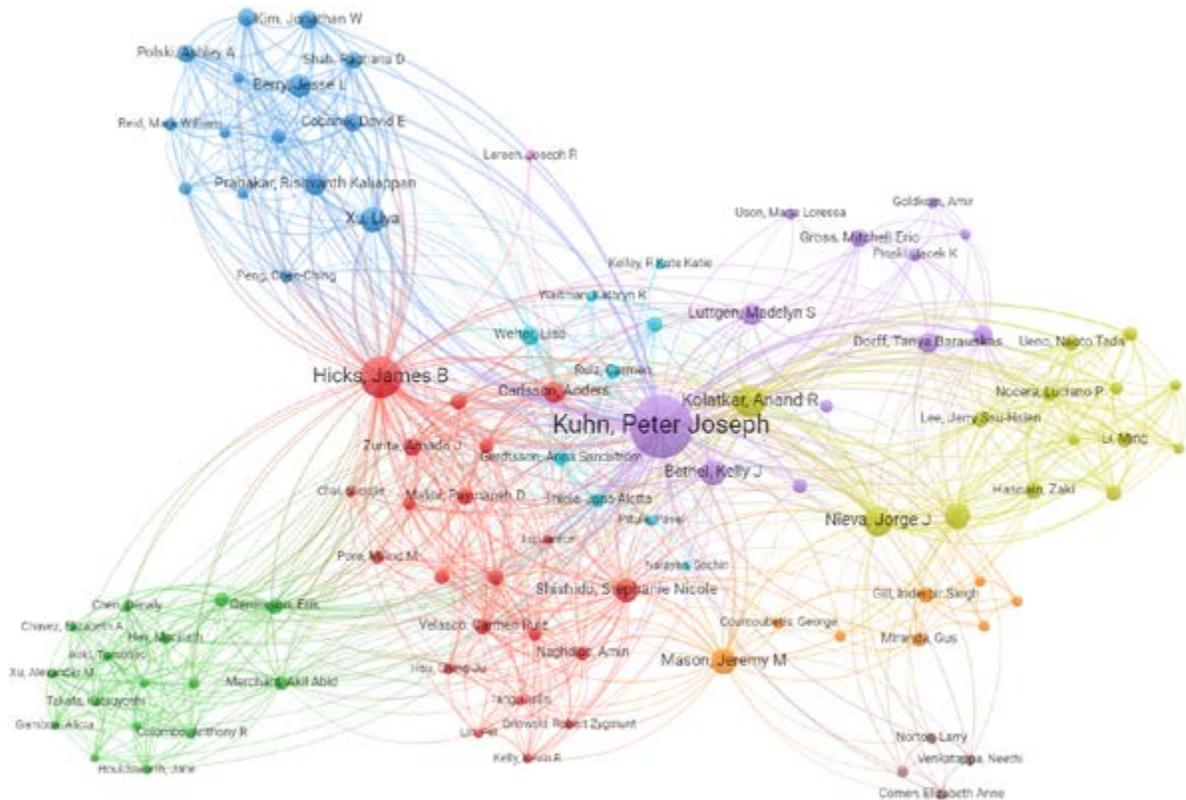
Gianluca Lazzi

Publications	Datasets	Grants	Patents
185	0	19	19
Citations		Funding amount	
4,053		USD 17.5 M	Clinical Trials
			0

Publications: Collaborations

Jeremy Mason

Publications	Datasets	Grants	Patents
29	6	0	0
Citations		Funding amount	
314		0	Clinical Trials
			0
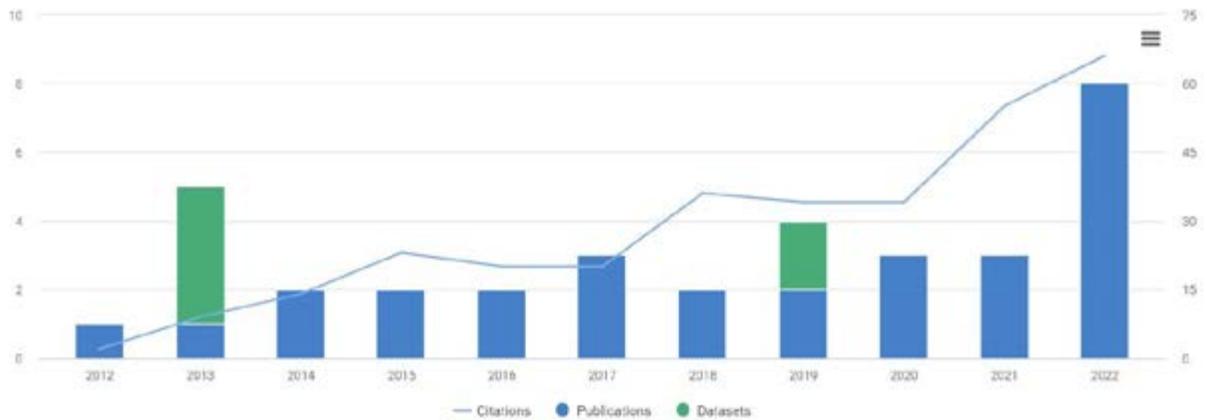
Publications: Collaborations
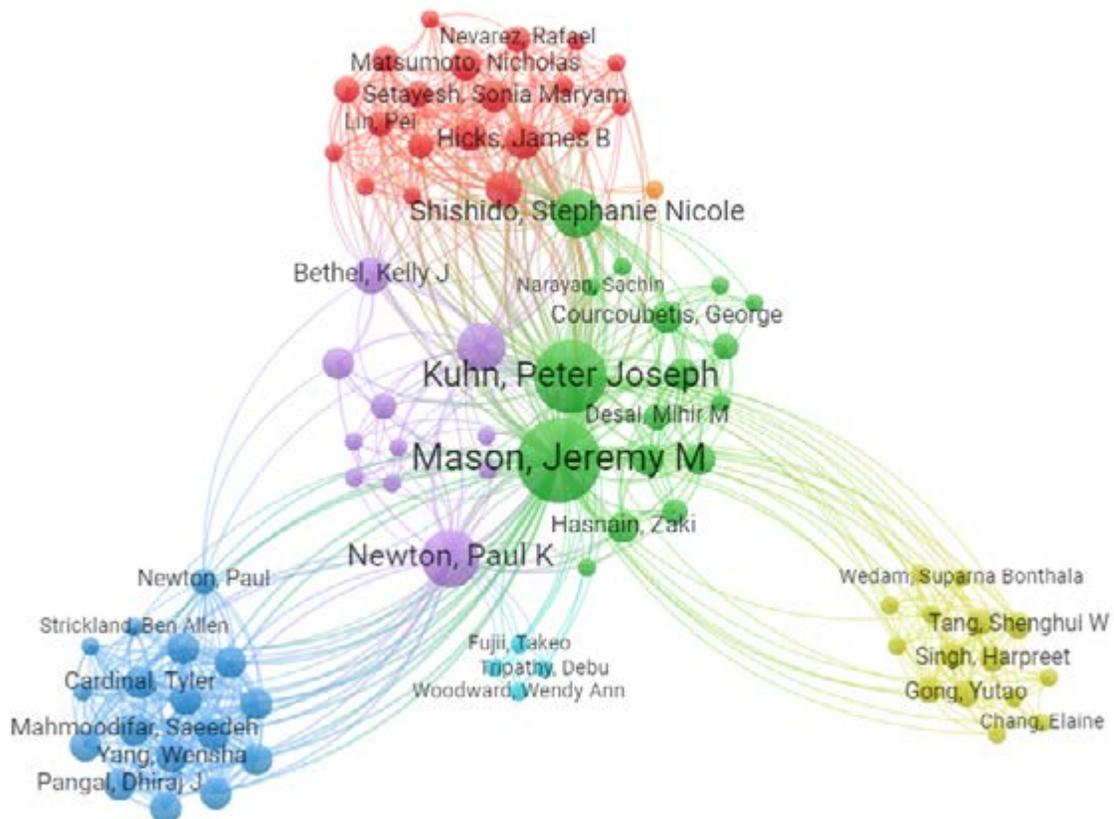
Ellis Meng

Publications	Datasets	Grants	Patents
167	1	21	93
Citations		Funding amount	Clinical Trials
4,256		USD 34.6 M	0

Publications: Collaborations

Shrikanth Narayanan

Publications	Datasets	Grants	Patents
989	47	26	25
Citations		Funding amount	
17,674		USD 30.2 M	Clinical Trials
			0

— Citations ● Publications ● Datasets ● Patents ● Active grants

Publications: Collaborations

Krishna Nayak

Publications	Datasets	Grants	Patents
214	8	11	32
Citations		Funding amount	
5,173		USD 10.5 M	
			Clinical Trials
			1

Publications: Collaborations

Niema Pahlevan

Publications	Datasets	Grants	Patents
55	15	5	76
Citations		Funding amount	
398		USD 1.6 M	Clinical Trials
			1

Publications: Collaborations

Jayakanth Ravichandran

Publications	Datasets	Grants	Patents
99	2	5	3
Citations		Funding amount	
1,992		USD 5.4 M	Clinical Trials
			0

Publications: Collaborations

Richard Roberts

Publications	Datasets	Grants	Patents
80	2	11	89
Citations		Funding amount	
5,751		USD 11.6 M	Clinical Trials
			0

Publications: Collaborations

Remo Rohs

Publications
103

Citations
7,112

Datasets
3

Grants
5

Funding amount
USD 5.5 M

Patents
0

Clinical Trials
0

— Citations ● Publications ● Datasets ● Active grants
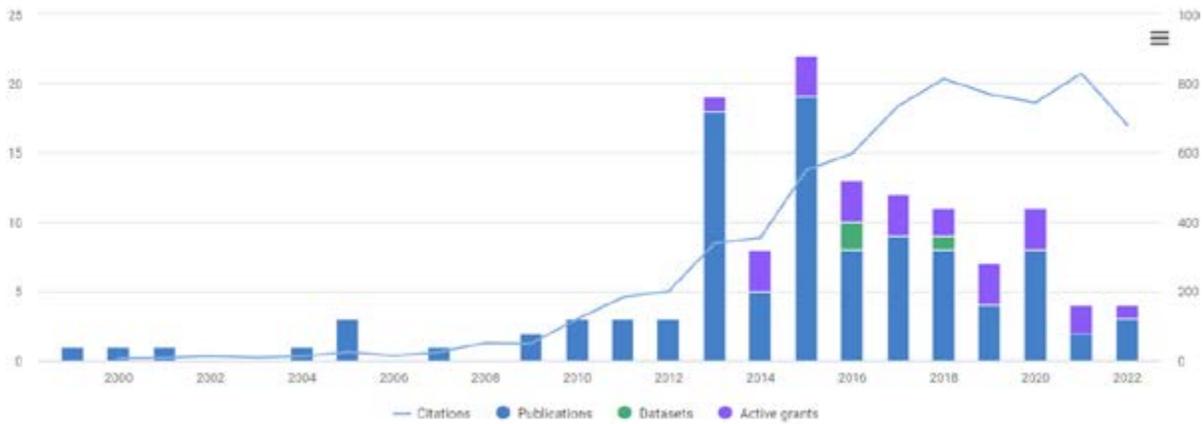
Publications: Collaborations
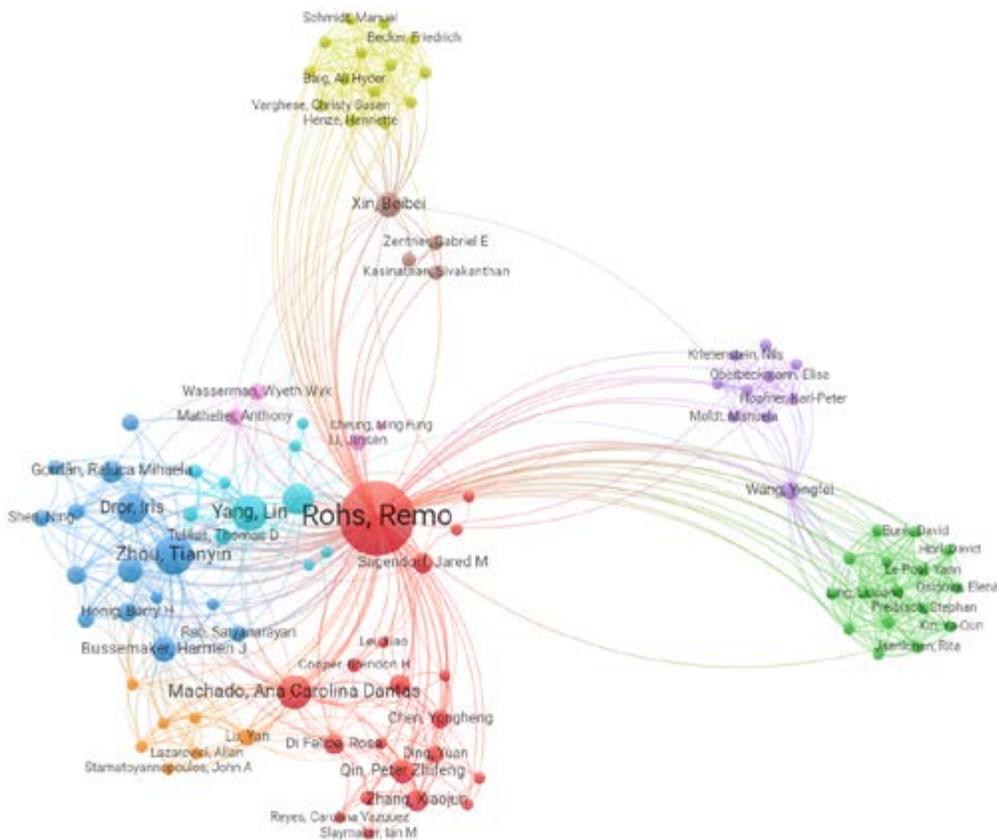
Maryam Shanechi

Publications: Collaborations

Terry Takahashi

Publications	Datasets	Grants	Patents
33	1	2	11
Citations		Funding amount	
1,150		USD 2.2 M	Clinical Trials
			0

Publications: Collaborations

Jennifer Treweek

Publications	Datasets	Grants	Patents
23	0	2	5
Citations		Funding amount	
1,615		USD 162 K	Clinical Trials
			0
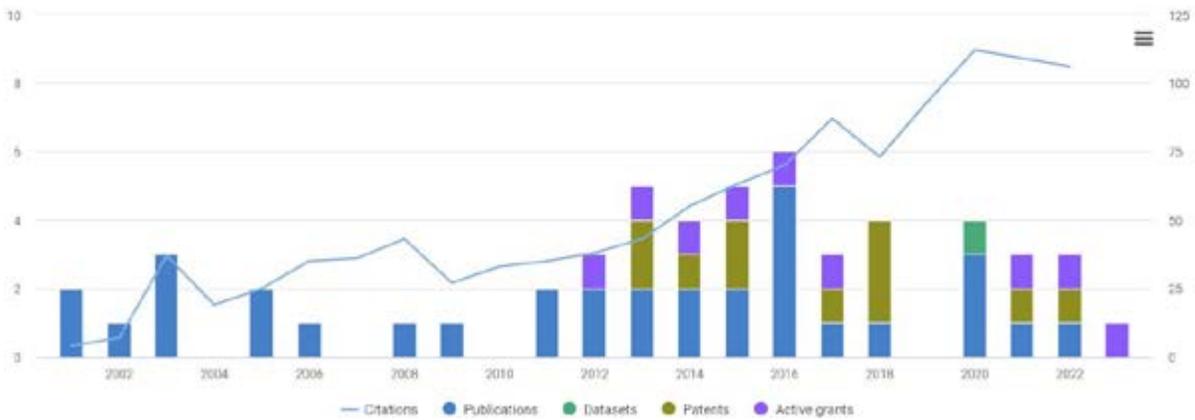
Publications: Collaborations
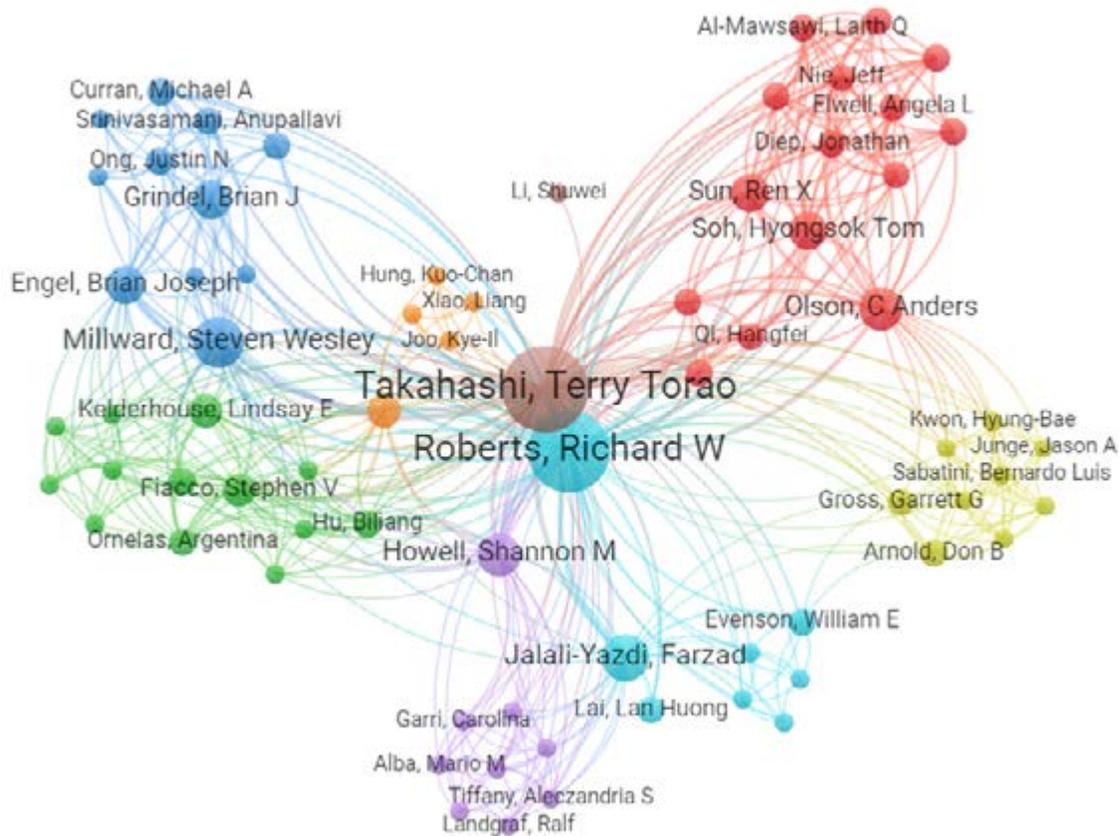
Kate White

Publications	Datasets	Grants	Patents
50	3	0	0
Citations		Funding amount	
1,376		0	Clinical Trials
			0
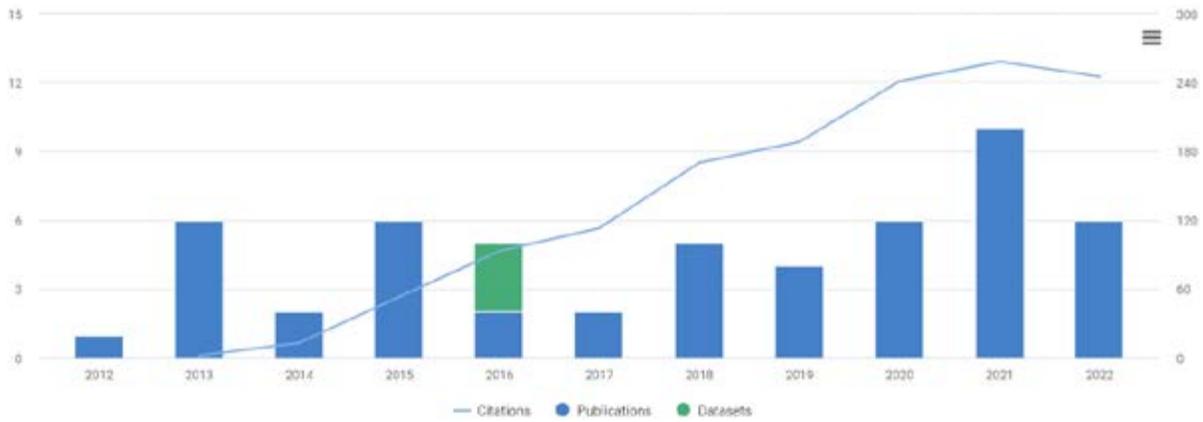
Publications: Collaborations
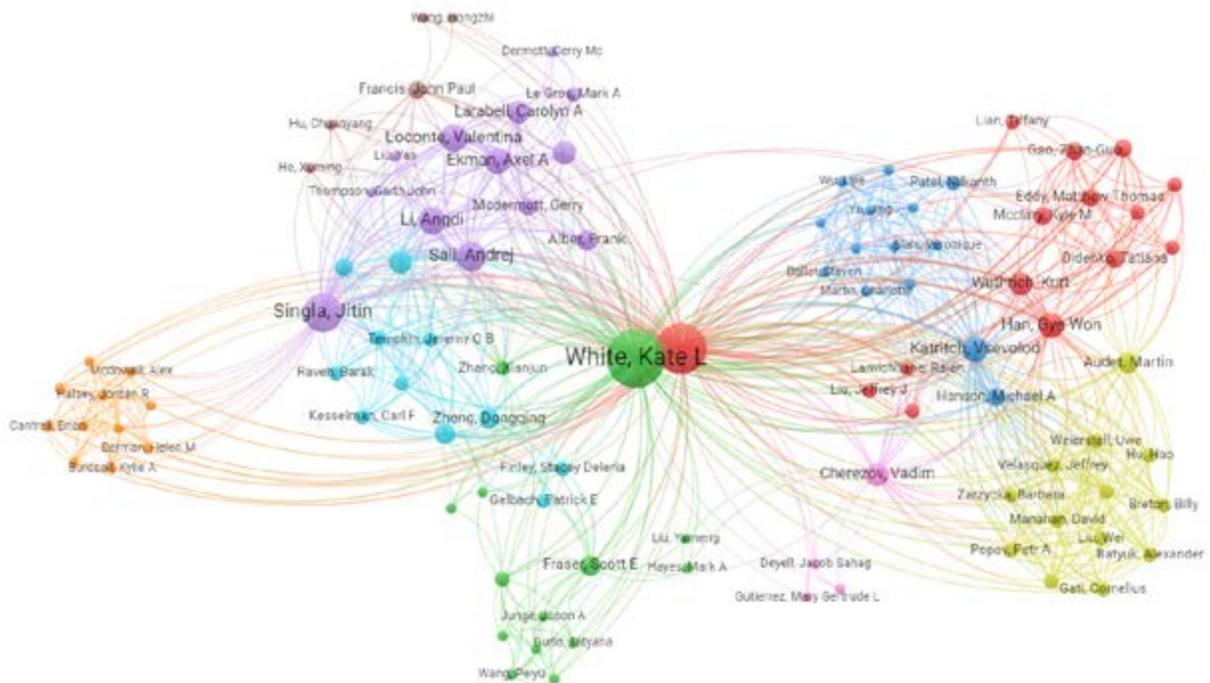
Cristina Zavaleta

Publications	Datasets	Grants	Patents
54	2	3	12
Citations		Funding amount	
5,129		USD 1.2 M	Clinical Trials
			0

Publications: Collaborations

Addendum: Journal Articles

Article

Structure-based design of bitopic ligands for the μ-opioid receptor

Check for updates

Abdelfattah Faouzi[1,10], Haoqing Wang[2,10], Saheem A. Zaidi[3,10], Jeffrey F. DiBerto[4,10], Tao Che[1,4,10], Qianhui Qu[2,5,10], Michael J. Robertson[2,5], Manish K. Madasu[1], Amal El Daibani[1], Balazs R. Varga[1], Tiffany Zhang[6], Claudia Ruiz[7], Shan Liu[8], Jin Xu[8], Kevin Appourchaux[1], Samuel T. Slocum[4], Shainnel O. Eans[9], Michael D. Cameron[7], Ream Al-Hasani[1], Ying Xian Pan[6,8], Bryan L. Roth[4], Jay P. McLaughlin[9], Georgios Skiniotis[2,5✉], Vsevolod Katritch[3✉], Brian K. Kobilka[2✉] & Susruta Majumdar[1✉]

Mu-opioid receptor (μOR) agonists such as fentanyl have long been used for pain management, but are considered a major public health concern owing to their adverse side effects, including lethal overdose[1]. Here, in an effort to design safer therapeutic agents, we report an approach targeting a conserved sodium ion-binding site[2] found in μOR[3] and many other class A G-protein-coupled receptors with bitopic fentanyl derivatives that are functionalized via a linker with a positively charged guanidino group. Cryo-electron microscopy structures of the most potent bitopic ligands in complex with μOR highlight the key interactions between the guanidine of the ligands and the key Asp[2.50] residue in the Na[+] site. Two bitopics (C5 and C6 guano) maintain nanomolar potency and high efficacy at G_i subtypes and show strongly reduced arrestin recruitment—one (C6 guano) also shows the lowest G_z efficacy among the panel of μOR agonists, including partial and biased morphinan and fentanyl analogues. In mice, C6 guano displayed μOR-dependent antinociception with attenuated adverse effects, supporting the μOR sodium ion-binding site as a potential target for the design of safer analgesics. In general, our study suggests that bitopic ligands that engage the sodium ion-binding pocket in class A G-protein-coupled receptors can be designed to control their efficacy and functional selectivity profiles for G_i, G_o and G_z subtypes and arrestins, thus modulating their in vivo pharmacology.

Opioid receptors, members of the class A family of G-protein-coupled receptors[4] (GPCRs), are key molecular targets in pain management. The challenge remains to engage specific subtypes of opioid receptor and to trigger a specific functional response that can produce in vivo analgesia without side effects such as tolerance, respiratory depression and addiction[5]. Some approaches to mitigate these adverse reactions[6] include the development of biased μOR agonists[7–11], peripherally restricted agonists[12], ligands targeting μOR splice variants[13], opioids with mixed actions at other subtypes[10,14,15] and compounds that bind to μOR only in acidic environments[16]. Most of these past approaches have targeted the orthosteric site, although positive allosteric modulators have recently been reported at μORs[17]. Here we aimed to target a distinct Na[+]-binding allosteric site, which is highly conserved in a vast majority of family A GPCRs[2] and is detected crystallographically in high-resolution inactive state structures[2,18] and biochemically in at least 26 diverse family A GPCRs.

Allosteric modulation of G_i and G_o GPCR subtypes by sodium was first observed in the 1970s at the μOR[19]. The presence of physiological NaCl concentrations was found to shift the receptor towards the inactive state, thus reducing agonist binding[2,20,21]. As recent structural studies reveal, the pocket undergoes marked conformational changes upon receptor activation and is critical for the modulation of signalling in family A GPCRs[22]. It has also been shown that the Na[+]-binding site residues serve as major 'efficacy switches'[18] that can bias the GPCR signalling to either G_i protein or β-arrestin-2 pathways. The polar cavity with the Na[+] site has also been characterized by molecular dynamics in inactive μOR[21,23], whereas the high-resolution structure of μOR bound to the agonist Bu72 (Protein Data Bank (PDB): 5C1M) revealed activation-related conformational changes in the cavity that prevent sodium binding but retain several water molecules[3] (Fig. 1a).

Moreover, this structure revealed a water-filled polar channel linking the orthosteric pocket with the Na[+] site, raising the possibility of

[1]Center for Clinical Pharmacology, University of Health Sciences and Pharmacy and Washington University School of Medicine, St Louis, MO, USA. [2]Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA, USA. [3]Department of Quantitative and Computational Biology, Department of Chemistry, Bridge Institute and Michelson Center for Convergent Bioscience, University of Southern California, Los Angeles, CA, USA. [4]Department of Pharmacology, University of North Carolina School of Medicine, Chapel Hill, NC, USA. [5]Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, USA. [6]Department of Neurology and Molecular Pharmacology, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [7]Department of Chemistry, Scripps Research, Jupiter, FL, USA. [8]Department of Anesthesiology, Rutgers New Jersey Medical School, Newark, NJ, USA. [9]Department of Pharmacodynamics, University of Florida, Gainesville, FL, USA. [10]These authors contributed equally: Abdelfattah Faouzi, Haoqing Wang, Saheem A. Zaidi, Jeffrey F. DiBerto, Tao Che, Qianhui Qu. ✉e-mail: yiorgo@stanford.edu; Katritch@usc.edu; kobilka@stanford.edu; susrutam@email.wustl.edu
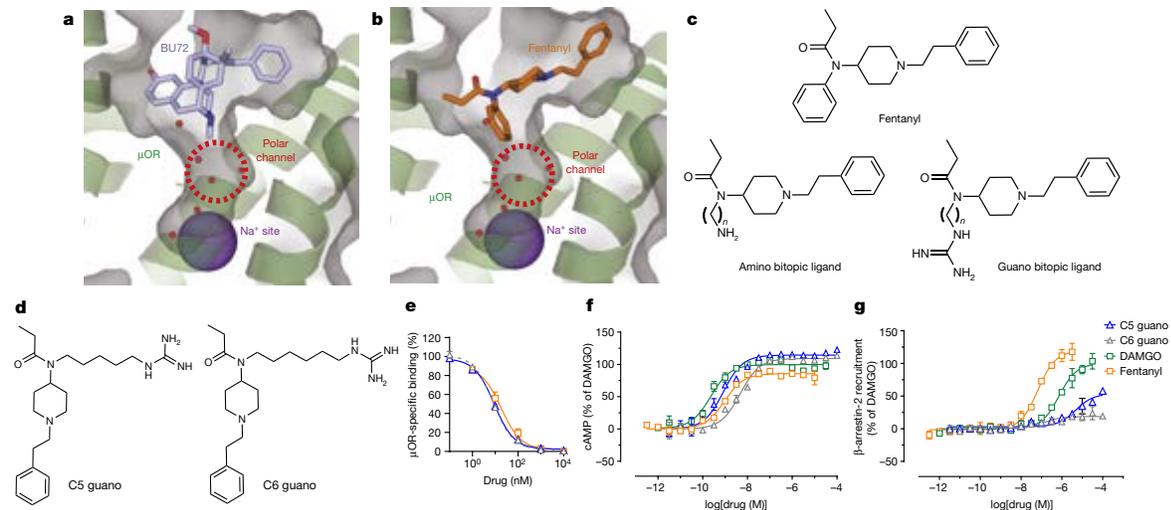
Article

Fig. 1 | Targeting the Na⁺ site with fentanyl-based bitopic ligands and characterization of lead compounds in binding, G-protein and arrestin signalling assays. a, Structure of Bu72 bound to μOR, showing the orthosteric site and the unoccupied Na⁺ site. The μOR structure is shown in green ribbon as well as transparent grey surface. **b**, Docking of fentanyl in μOR, showing the orthosteric site and the unoccupied Na⁺ site. **c**, Chemical structures of fentanyl and designed bitopic ligands. **d**, Chemical structures of the lead fentanyl guano bitopic ligands. **e**, Binding affinities at the μOR. Lead bitopic ligands (guano fentanyls) were characterized in binding assays in CHO cells expressing μOR

using [¹²⁵I]IBNtxA as radioligand. C5 guano and C6 guano had similar affinity to fentanyl. Data are mean ± s.e.m. (n = 3 experiments each done in triplicate). **f**,**g**, cAMP inhibition (**f**) and Tango assay for β-arrestin-2 recruitment on μOR (**g**) with the bitopic ligands. The guano bitopic derivatives show high G-protein agonism with poor recruitment of β-arrestin-2; C5 guano was the most active relative to DAMGO. Data are mean ± s.e.m. (n = 3 experiments each done in duplicate). Extended Data Table 1 shows values for all panels and data for all bitopic ligands.

developing ligands that interact directly with the Na⁺ site. The high conservation and key functional role of this allosteric pocket suggest that it is suitable as a target for allosteric modulators and bitopic ligands with unique properties for some family A GPCRs, including the μOR. To address this hypothesis, we used a structure-based approach and designed a series of bitopic ligands on the fentanyl scaffold. These bitopic ligands target both the orthosteric pocket of the μOR and extend into the polar channel towards the Na⁺ site[24] (Fig. 1b,c). The binding pose of two fentanyl bitopic ligands was confirmed by structure determination using cryo-electron microscopy (cryo-EM). Functional characterization of these ligands revealed that the extension of a positively charged guanidine group into the Na⁺ site resulted in a near complete loss of arrestin recruitment, and, surprisingly, substantial changes in the relative potency and efficacy for the six G_i, G_o and G_z subtypes. Finally, the lead bitopic ligand, C6 guano, showed antinociception with reduced side effects compared with classical μOR agonists such as morphine or the lethal agonist fentanyl. Owing to the high conservation of the Na⁺ binding pocket, similar bitopic ligand designs may be suitable for effective control of functional selectivity in other class A GPCRs.

Bitopic μOR ligands based on the fentanyl scaffold

Fentanyl is a synthetic μOR-selective agonist that is 100-fold more potent as an analgesic than morphine[11,25]. Unlike morphine, fentanyl also is highly efficacious for inducing arrestin translocation and is a full agonist for G_i-mediated signalling, whereas morphine is a partial agonist. Fentanyl also acts rapidly and is commonly used to manage postoperative and severe cancer-related pain, and is known to bind specifically to the orthosteric site of μOR—but in the USA, its use has led to around 60,000 deaths due to overdose[26].

Here we predicted the binding poses of fentanyl on the basis of the crystal structure of the active state μOR with the N-terminus truncated

at residue M65[3] (Fig. 1b), and computationally designed and synthesized a small library of fentanyl derivatives designed to extend through the polar channel below the orthosteric pocket to the Na⁺-binding pocket (Extended Data Fig. 1). Although alternative models for fentanyl binding have been suggested[27,28], the docking model used here has been validated by cryo-EM with the fentanyl analogue lofentanil[29], as well as by cryo-EM complexes with the designed bitopic ligand reported here. The models predicted the distance between the amide nitrogen of fentanyl and the carboxy group of the D114²·⁵⁰ residue of the allosteric Na⁺-binding site to be approximately 13 Å (superscripts indicate Ballesteros–Weinstein numbering for GPCRs[30]). In order to engage the allosteric Na⁺-binding site, we replaced the classical fentanyl benzene ring with an aliphatic chain linker (C$_n$, where n = 3, 5, 6, 7, 9 or 11) connected to a positively charged 'warhead', an amine or guanidine (guano) group (Fig. 1c). The bitopic compound with a C7 linker (C7 guano) was predicted to be optimal for the formation of a direct salt bridge between the amino warhead and D114²·⁵⁰ carboxylic acid, and the C6 linker (C6 guano) was predicted to be optimal for the interaction for the guanidine warhead. The syntheses of amino and guano bitopic ligands[31] on the fentanyl scaffold are presented in Supplementary Fig. 1.

Functional assessment of fentanyl-based scaffold

We first evaluated the binding affinities of fentanyl bitopic ligands on μOR using a competitive radioligand assay with labelled 3′-iodobenzoyl-6β-naltrexamide[13] ([¹²⁵I]IBNtxA) (Fig. 1e and Extended Data Table 1). The binding data clearly indicate a difference between guano- and amino-fentanyl bitopic ligands. The amino-fentanyl compounds showed reduced affinity towards μOR, suggesting that the amino warhead does not allow efficient interaction with the polar and negatively charged residues predicted to line the sodium binding pocket (D114²·⁵⁰, N150³·³⁵, N328⁷·⁴⁵, S329⁷·⁴⁶ and S154³·³⁹). Furthermore, the extension of the aliphatic linker did not seem to substantially affect the binding of
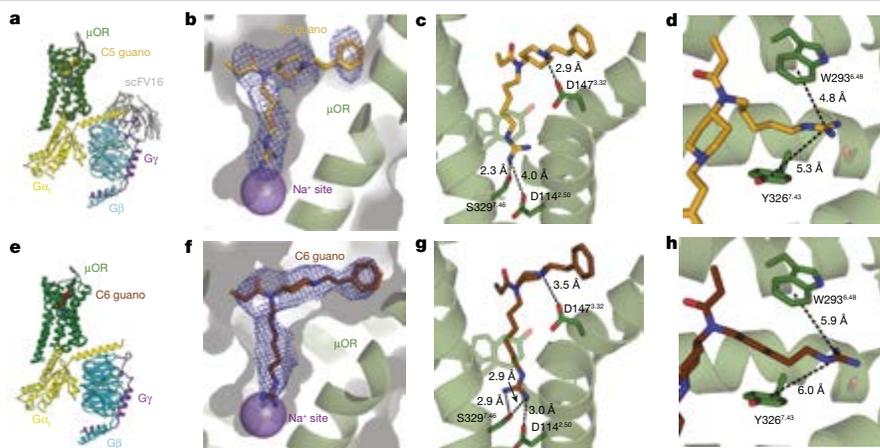
Fig. 2 | Structures of bitopic ligands bound to µOR. a,e, Cryo-EM structure of C5 guano (**a**) and C6 guano (**e**) bound to µOR–G$_i$–scFv16 complex coloured by subunit. **b,f,** View of the C5 guano (**b**) and C6 guano (**f**) cryo-EM density. The cryo-EM maps are contoured at 5.0σ. The functional groups of the bitopic ligands target site 2 above the Na$^+$-binding site, as expected. **c,g,** Salt-bridge interactions between C5 guano (**c**) or C6 guano (**g**) and residues in the µOR pocket. **d,h,** Cation–π interactions between C5 guano (**d**) or C6 guano (**h**) and residues in the µOR pocket.

such molecules, with (inhibition constant) K_i values of 281 and 369 nM for C9 amino and C7 amino ligands, respectively.

By contrast, all guanidino derivatives displayed moderate to high affinity towards µOR with K_i values ranging from 590 nM for the C11-guano derivative to 4.6 nM for C5 guano and 4.1 nM for C6 guano, which represents a slight improvement over fentanyl binding (K_i = 9 nM). The guanidine moiety appears to provide a better fit and additional hydrogen bond donors and/or acceptors, potentially enhancing interaction in the polar cavity of µOR, with K_i values differing by up to 1,000-fold compared with their amino counterparts. Structure–activity relationship analysis showed that whereas the optimal affinity was obtained with C5 guano and C6 guano, C3 guano was too short to reach the Na$^+$ pocket and showed reduced binding affinity (K_i = 77 nM). Further extension of the bitopic linker beyond C6 had a negative effect on the affinity. C7 guano showed increased K_i values, probably owing to steric clash because of inadequate space in the polar pocket for the extended linker.

To assess the functional properties of the bitopic ligands on µOR, we used the Tango assay for β-arrestin-2 recruitment and cAMP assays for G$_i$ protein activation (Fig. 1f,g and Extended Data Table 1). As expected from their low affinity, the amino fentanyls displayed only weak stimulation of the G$_i$ pathway, with half-maximal effective concentration (EC$_{50}$) values all above 293 nM and no measurable β-arrestin-2 recruitment at the highest tested drug concentrations of 10 µM for the majority of analogues. Among guanidino bitopic ligands, the highest affinity C5 guano ligand demonstrated full agonistic activity in cAMP assays (maximum G-protein activation (E_{max}) = 114%) with subnanomolar potency (EC$_{50}$ = 0.8 nM), in the same range as the standard full agonist DAMGO (EC$_{50}$ = 0.9 nM). Following the same trend as affinity, the potency of guanidine derivatives was inversely correlated with the aliphatic linker chain length from C6 guano (EC$_{50}$ = 4.7 nM) to C11 guano (EC$_{50}$ = 1017 nM).

Of note, the bitopic guanidine-containing derivatives with the strongest affinity and potency in cAMP assays did not recruit β-arrestin-2 efficiently. For example, C5 guano in Tango assays had EC$_{50}$ = 4,710 nM and E_{max} = 54%, whereas for DAMGO, EC$_{50}$ = 723 nM and E_{max} = 100%, and for fentanyl, EC$_{50}$ = 66 nM and E_{max} = 119 %. For C6 and C7 derivatives in Tango assays, E_{max} was less than 20% relative to DAMGO, and in the cAMP assay, E_{max} was 109% and 107% relative to DAMGO, respectively. Thus, the designed extensions of the fentanyl scaffold into the polar pocket below the orthosteric site markedly reduced the β-arrestin-2

recruitment capacity of fentanyl[11], which is considered one of the strongest β-arrestin-2 recruiting opiates.

Structural validation of bitopic ligand design

Given the pharmacological data showing that C5 guano and C6 guano are the most potent bitopic ligands with respect to cAMP inhibition with reduced β-arrestin-2 activity, we used cryo-EM to determine the structures of the µOR–G$_{i1}$ complex bound to C5 guano and C6 guano at 3.2 Å and 3.3 Å global resolution, respectively (Fig. 2, Extended Data Figs. 2 and 3 and Extended Data Table 2). The bitopic ligand-bound µOR structures reveal the expected binding pose of the ligands, with their fentanyl scaffolds overlapping with the fentanyl model in the orthosteric pocket and the guanidine extension protruding towards the sodium site (Fig. 2b,c,f,g). The conformation of the bitopic ligand-bound receptor is very similar to the high-resolution structure of BU72-bound active µOR (PDB: 5C1M) with all-atom root-mean-squared deviation (r.m.s.d.) being 0.957 Å for C5 guano and 1.189 Å for C6 guano, for residues lining the orthosteric pocket (Extended Data Fig. 4). Unlike C5 guano and C6 guano, BU72 does not extend into the polar pocket and the all-atom r.m.s.d. of the residues predicted to form the Na$^+$-binding pocket and polar channel is 0.884 Å for C5 guano and 1.056 Å for C6 guano (Extended Data Fig. 4b,d).

In both structures, we observe direct interaction between the orthosteric residue D147[3.32] and the piperidine ring nitrogen atom, which is consistent with the docked fentanyl model (Fig. 2c,g). This confirms that C5 and C6 guano bitopic ligands maintain the same specific interaction to the orthosteric pocket as fentanyl.

The structure also reveals that the shorter C5 linker affords only limited interaction (distance 4.0 Å) between the basic guanidine warhead of C5 guano and the major anchor of the Na$^+$ binding site, the acidic D114[2.50] side chain (Fig. 2c). Nevertheless, it is possible that an interaction between the warhead and the aspartate is dynamically mediated by water molecules that are not resolved in the 3.2 Å resolution cryo-EM map. The lack of direct contact with D114[2.50] can be compensated by a strong hydrogen-bonded local interaction of the guanidine warhead with S329[7.46], which is another key residue of the Na$^+$-binding site (Fig. 2c). The structure of C6 guano-bound µOR has subtle localized differences compared with C5 guano-bound µOR. In the C6 guano-bound structure, the distance between D114[2.50] and
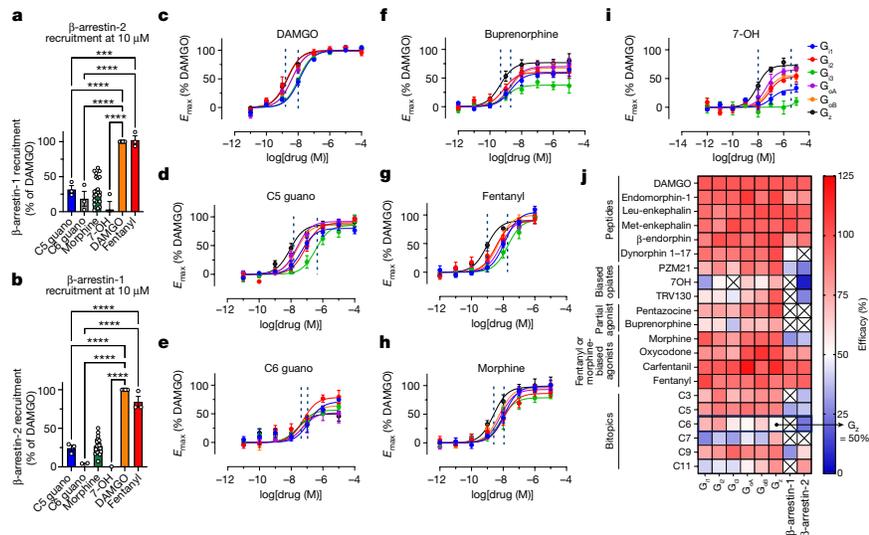
Article

Fig. 3 | Profiling of C5 guano, C6 guano and μOR using TRUPATH Gαβγ biosensors and β-arrestin-1 and β-arrestin-2 efficacy. a,b, BRET assays for β-arrestin-2 (**a**) and β-arrestin-1 (**b**) recruitment in the presence of 10 μM C5 guano, C6 guano, morphine, fentanyl, buprenorphine, 7-OH and DAMGO (all relative to DAMGO). **a,** C5 guano and C6 guano showed significantly reduced β-arrestin-2 recruitment compared with DAMGO. Minimum, median and maximum values−C5 guano: 23.9%, 29.0%, 42,5%; C6 guano: 2.7%, 14.6%, 38.4%; morphine: −6.0%, 25.8%, 62.6%; 7-OH: −13.1%, −2.4%, 25.2%; DAMGO: 100%, 100%, 100%; fentanyl: 93.5%, 99.7%, 113.8%. **b,** Similarly, C5 guano and C6 guano showed significantly reduced β-arrestin-1 recruitment compared with DAMGO. Minimum, median and median values−C5 guano: 17.0%, 26.6%, 29.4%; C6 guano: −7.9%, 4.2%, 4.8%; morphine: 7.2%, 25.6%, 50.2%; 7-OH: −8.0%, −1.4%, 0.3%;

DAMGO: 100%, 100%, 100%; fentanyl: 77.4%, 77.4%, 98.4. Data are mean ± s.e.m. (*n* = three experiments each done in duplicate). Primary statistics for **a,b** are provided in Supplementary Table 2. **c–i,** Gα-subtype selectivity using TRUPATH on μOR for DAMGO (**c**), C5 guano (**d**), C6 guano (**e**), buprenorphine (**f**), fentanyl (**g**), morphine (**h**) and 7-OH (**i**). C6 guano showed distinct potencies and efficacies for all six Gα-protein subtypes (G_{i1}, G_{i2}, G_{i3}, G_{oA}, G_{oB} and G_z) with the lowest efficacy for the G_z subtype. **j,** Efficacy heat map for opioid peptides, biased agonists, partial agonists, morphine or fentanyl template agonists and bitopic ligands using TRUPATH and β-arrestin-1 and β-arrestin-2 activity. Raw curves and values for all compounds are presented in Extended Data Fig. 6 and Extended Data Tables 2 and 3. Data are mean ± s.e.m. (*n* = three experiments each done in duplicate).

the guanidine warhead was 3 Å, enabling the formation of a direct salt bridge. Besides interacting with the polar residues of the Na⁺ binding site, the positive charge of the guanidine warhead might also form weak cation–π interactions with the aromatic residues located at the polar channel[32] (Fig. 2d,h), further stabilizing the binding pose of guano bitopic ligands with appropriate aliphatic linker sizes.

Molecular dynamics of the μOR complexes

The results presented above demonstrate that the unique binding interactions observed for C5 guano and C6 guano bitopic ligands with μOR (Fig. 2) lead to signalling profiles that are very distinct from their orthosteric templates. To obtain further conformational insights into the ligand–receptor interactions, we analysed the dynamics of the direct and water-mediated interactions of C5 guano and C6 guano and other derivatives in the active state μOR. We performed 10 independent molecular dynamics runs of approximately 1-μs duration for each complex in a lipid bilayer membrane (Extended Data Fig. 5). The initial structure for the C5 guano molecular dynamics simulation was derived from the corresponding cryo-EM structure, whereas the initial C6 guano complex was obtained by docking of the ligand into the structure of μOR defined by the C5 guano complex. The molecular dynamics trajectories were used to calculate (1) the distance between the ligand piperidine nitrogen atom and carboxylate oxygens of D147[3.32], (2) the distance between guano nitrogen atoms of the ligand and carboxylate oxygens of D114[2.50], and (3) the number of waters bridging the guanidine group and carboxylate oxygens of D114[2.50] (Extended Data Fig. 5f,g). The conformations of the fentanyl core in the orthosteric pocket were stable for all ligands, with more than 96% of molecular dynamics frames

maintaining either direct salt bridge (more than 85% of frames, defined as an N–O distance of less than 3.5 Å) or water-mediated interactions between piperidine nitrogen and carboxylate oxygens of D147[3.32]. The interactions between guanidine and D114[2.50] were also maintained in at least 90% of the frames, however, with different ratios of direct salt bridges and water-mediated interactions. For C5 guano, 33% of the trajectory frames had 1 or 2 water molecules bridging guanidine group and D114[2.50], whereas 57% of the frames showed direct salt-bridge interactions between C5 guano and D114[2.50], with a relatively rapid exchange between these interaction modes. For C6 guano, this direct salt-bridge bonding to D114[2.50] increased to 71% of all frames, with an additional 23% of indirect interactions, in line with a more optimal distance match predicted for the C6 linker. Our observation of a vast majority (more than 90%) of the molecular dynamics conformations maintaining either a water-mediated or a direct interaction between C5 guano and D114[2.50] of the Na⁺-binding pocket of μOR suggests the importance of these interactions for the ligand affinity and functional properties.

Intrinsic efficacy compared with classical agonists

We next pharmacologically assessed our guano bitopic ligands using bioluminescence resonance energy transfer (BRET)-based assays for measuring arrestin recruitment and G-protein heterotrimer dissociation (TRUPATH), including at both arrestin isoforms (β-arrestin-1 and β-arrestin-2) and at all Gα subtypes (G_{i1}, G_{i2}, G_{i3}, G_{oA}, G_{oB} and G_z), through which μOR signals[25,33]. Specifically, we compared their signalling profiles with those of several other μOR ligands, including chemically and functionally diverse endogenous and exogenous agonists (Fig. 3, Extended Data Fig. 6 and Extended Data Tables 3 and 4). For brevity,

we focus here on our two most potent compounds, C5 guano and C6 guano, in comparison with selected other ligands with pharmacological reference to the synthetic peptide agonist DAMGO.

Previous efforts to develop opioid drugs with reduced side effects have been based on the hypothesis that β-arrestin-2 mediates deleterious responses such as respiratory depression[7,11,34]. The β-arrestin-2 recruitment profile of bitopic ligands was similar to the Tango assay. C5 guano (E_{max} = 38%) and C6 guano (E_{max} = 22%) showed progressively reduced β-arrestin-2 efficacy compared with fentanyl (E_{max} = 97%) but similar to morphine (E_{max} = 27%) (Fig. 3a). The G-protein-biased compounds, such as PZM21, 7-OH and TRV130, weakly recruited β-arrestin-2 (E_{max} = 24%, no response and 26%, respectively). Of note, these G-protein-biased compounds, for which we observed no or reduced β-arrestin-2 recruitment over the tested concentration range, have shown adverse effects (including respiratory depression) in certain preclinical animal models[35-37] and in human clinical trials[38], suggesting that low or even no β-arrestin-2-independent signalling is insufficient for at least certain unwanted physiological responses[6,9,39].

We also examined β-arrestin-1 recruitment. C5 guano showed 32% efficacy in β-arrestin-1 recruitment assays—compared with 92% for fentanyl and 27% for morphine—whereas other bitopic ligands such as C6 guano showed no measurable β-arrestin-1 efficacy (Fig. 3b).

It has been proposed recently that low intrinsic G-protein efficacy, rather than abolished β-arrestin-2 activity, more favourably results in reduced side effects[35]. This suggestion was made on the basis of BRET-based measurements of heterotrimeric G-protein dissociation, which displays lower amplification and sensitivity to receptor reserve than conventional functional assays, and thus more accurately reflects measurements of ligand intrinsic efficacy[33]. As μOR signals through all G_i, G_o and G_z family members, and this previous study only investigated one G_{i2} subtype, we sought to more thoroughly characterize the concept with both existing and our novel ligands using the TRUPATH platform. The ligands differed in the efficacy and order of potency for specific G_i, G_o and G_z subtypes, as well as in the range of potencies from the most- to least-sensitive G-protein isoform (dashed lines in Fig. 3c–i indicate the range of potencies; Extended Data Fig. 6, Extended Data Tables 3 and 4 and Supplementary Fig. 2). For DAMGO, the potency order is (G_z ≈ G_{i2} > G_{oA} ≈ G_{oB} > G_{i3} ≈ G_{i1}) with the EC_{50} ranging from 2.1 nM for G_z to 12.4 nM for G_{i1}. Fentanyl had a similar potency order (G_z > G_{i2} > G_{oA} ≈ G_{oB} > G_{i1} > G_{i3}), whereas the EC_{50} range was greater, 0.9 nM for G_z to 19 nM for G_{i3}. Morphine had a lower efficacy than DAMGO at all G_i and G_o subtypes, with the lowest efficacy at G_{i3} (79%). The distribution of EC_{50} values for morphine was similar to that for DAMGO (G_z > G_{oB} ≈ G_{oA} ≈ G_{i3} > G_{i2} ≈ G_{i1}), ranging from 2.4 nM for G_z to 11 nM for G_{i1}.

Similarly diverse profiles were observed with most exogenous opioids, whereas endogenous opioids were full or nearly full agonists at all transducers tested (an exception being Dyn1-17 at the arrestins). For C5 guano and C6 guano, we observed similarly unique signalling profiles, including between these compounds, despite their chemically similar scaffolds. For example, C5 guano displayed a wide range of potencies (lowest and highest EC_{50}: G_{i3}, 350 nM and G_z, 8.8 nM) but a narrow range of efficacies (lowest and highest E_{max}: G_{i1}, 80% and G_{oA}, 92%), whereas C6 guano displayed a narrow range of potencies (lowest and highest EC_{50}: G_{i1}, 133 nM and G_z, 25n M) but a wide range of efficacies (lowest and highest E_{max}: G_z, 50% and G_{i2}, 79%), with overall lower efficacy. Of note, we observed that C6 guano displayed a lower efficacy at G_z both relative to other G proteins and in terms of efficacy relative to that of all other agonists (Fig. 3j). Notably, this includes the partial agonists buprenorphine and pentazocine (E_{max}, G_z = 78% and 103%) as well as biased agonists such as 7-OH, TRV130 and PZM21 (E_{max}, G_z = 73%, 98% and 98% respectively).

In summary, C6 guano shows marginal decreases in G-protein potency and lower intrinsic efficacy relative to classical μOR agonists such as DAMGO, fentanyl, morphine and oxycodone, and had the lowest efficacy for G_z. C6 guano showed diminished arrestin recruitment relative to DAMGO, other opioid peptides, fentanyl and oxycodone, but similar arrestin recruitment to morphine.

Selectivity across species and other GPCRs

Although there is high species homology between human μOR and mouse μOR, given that the lead bitopic ligand was to be profiled in mice, we screened C5 guano and C6 guano and several other μOR agonists for binding and G_{i1} and β-arrestin-2 activity (Supplementary Table 1). The binding affinity for C6 guano at both human and mouse μOR was similar—that is, 1.2 and 1.08 nM, respectively (Supplementary Table 1a). Differences in G_{i1} potency and efficacy across drugs across species were marginal. The EC_{50} of C6 guano at G_{i1} was 132 nM for human μOR and 39 nM for mouse μOR, and the E_{max} was 71% and 79%, respectively. Crucially, C6 guano retains partial agonism and poor β-arrestin-2 recruitment at mouse μOR (Supplementary Table 1b).

C5 guano and C6 guano were counter-screened across around 45 targets using competitive radioligand binding assays through the Psychoactive Drug Screening Program at the National Institute of Mental Health[40]. In this assay platform, the affinity at μOR was 6.1 and 2.7 nM for C5 guano and C6 guano, respectively, and there was very little off-target binding (Supplementary Table 1c). The highest off-target affinities were found at the κ-opioid receptor (κOR) (224 and 259 nM for C5 and C6 guano bitopic ligands, respectively) and at the δ-opioid receptor (δOR) (1,530 and 312 nM, respectively), showing around 100-fold selectivity for C6 guano. The nearest non-opioid target for C5 guano was the histamine H1 receptor, over which it exhibited 45-fold selectivity; C6 guano exhibited more 400-fold selectivity over H_1 and the $α_{1A}$ adrenoreceptor. When screened in BRET-based G_{i1} signalling assays at κOR and δOR, both C5 and C6 guano showed reasonable selectivity for μOR (Extended Data Table 1 and Supplementary Figs. 2 and 3a–c).

In vivo pharmacology of C6 guano

Given the distinct differences in Gα-subtype signalling displayed by C6 guano compared with the other agonists, we further examined the pharmacology of this analogue in vivo using mouse models. The antinociceptive effect of C6 guano was evaluated in vivo in mice using a standard 55 °C warm water tail-withdrawal assay[10,15], with the compound administered supraspinally (intracerebroventricular (ICV)), since it showed no systemic activity when administered intraperitoneally, probably because the two positive charges result in suboptimal blood–brain barrier penetration. The median antinociceptive dose (ED_{50}) by ICV (and 95% confidence interval) for C6 guano was 18.77 nmol (5.49–55.54 nmol) (Fig. 4a), slightly higher than the ED_{50} of morphine, 6.6 nmol (4.4–8.43 nmol) (Supplementary Fig. 4a). A ceiling effect of antinociception was observed with C6 guano with increasing dosage from 30 to 300 nmol. C6 guano antinociception was found to be significantly attenuated in μOR-knockout mice (Fig. 4b). Before further analysis, we tested its stability in brain homogenates and carried out brain exposure studies. C6 guano showed high metabolic stability (Supplementary Fig. 4b). When assessed in pharmacokinetics assays at a 100 nmol ICV dose, C6 guano showed decent exposure at all tested time points. At the peak antinociceptive time point of 20 min, C6 guano showed 450 fold higher brain exposure than the ligand's agonistic potency at mouse μOR (Supplementary Fig. 4c). These results suggest that C6 guano is stable and has desirable μOR occupancy for in vivo activity. C6 guano was further characterized in detail for potential adverse effects. Locomotor effects were evaluated using the Comprehensive Lab Animal Monitoring System (CLAMS) assay[10,25]. Ambulations induced by 300 nmol C6 guano did not differ significantly from vehicle, in contrast to morphine, which showed hyperlocomotion at 100 and 300 nmol doses (5× and 15× of the ED_{50}, respectively) (Fig. 4c). We also measured breath rates in the same CLAMS assay at 100 and 300 nmol doses. Treatment with C6
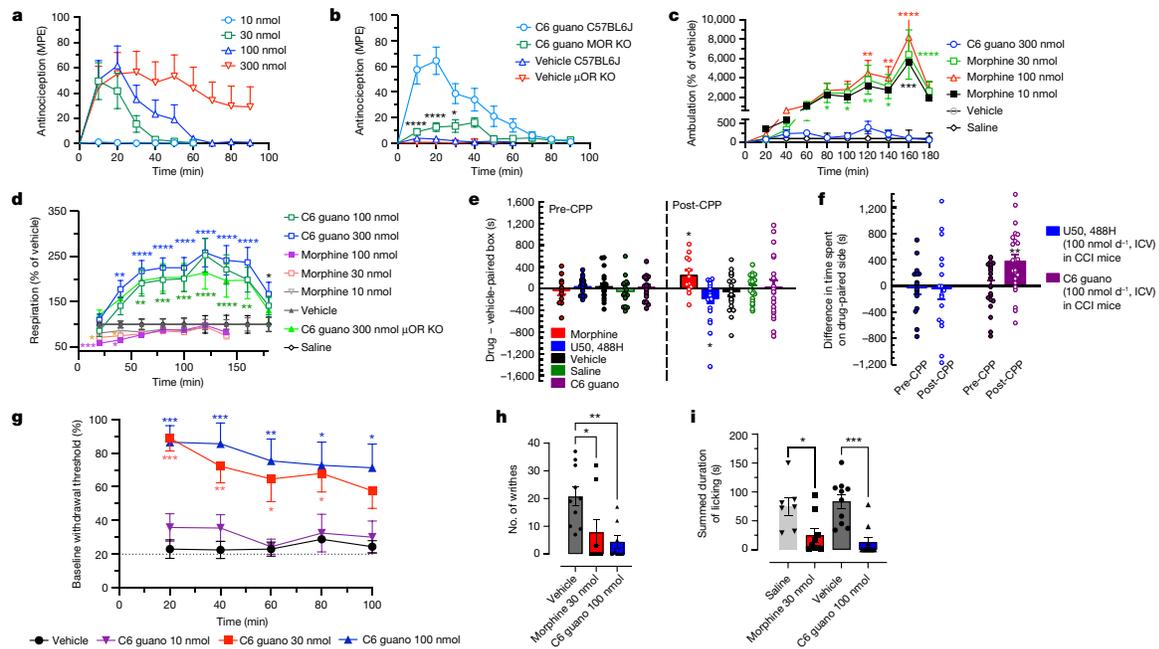
Article

Fig. 4 | C6 guano exhibits μOR-mediated antinociception without CPP, CPA or hyperlocomotor effects. a, Antinociceptive time course: C57BL/6J mice were administered C6 guano by ICV injection and antinociception was measured using the 55 °C tail-withdrawal assay. Data are shown as mean percentage antinociception (MPE) ± s.e.m. An ED$_{50}$ (with 95% confidence interval) of 18.77 nmol (5.49–55.54 nmol) was calculated for C6 guano. **b**, Antinociception by 100 nmol ICV C6 guano was attenuated in μOR-knockout (KO) mice compared with wild-type (C57BL/6J) or vehicle-treated (veh) mice. Data are MPE ± s.e.m. **c**,**d**, Locomotor effects and respiratory depression. C57BL/6J mice were administered either saline, vehicle, morphine (10, 30 or 100 nmol ICV), C6 guano (100 or 300 nmol ICV) and ambulation (**c**) or number of breaths (**d**) was measured every minute and averaged in 20-min bins. Data are presented as percentage of vehicle response ± s.e.m. **c**, Morphine treatment resulted in hyperlocomotion, whereas the C6 guano (300 nmol ICV) effect was not significantly different from ICV vehicle at any time point. **d**, Morphine did not result in any significant decrease in respiration rate at 10 nmol, but caused respiratory depression at 30 nmol and 100 nmol doses. C6 guano caused increased respiration rate at 100 nmol and 300 nmol doses compared with vehicle. There were no significant differences between wild-type and μOR-knockout mice treated with 300 nmol ICV C6 guano. **e**, C6 guano (100 nmol ICV) treatment did not result in CPP or CPA changes relative to vehicle, whereas morphine (30 nmol ICV) resulted in CPP and U50,488H (100 nmol ICV) resulted in CPA. Minimum, median and maximum values—U50,488H pre-conditioning (pre-CPP): −207 s, 52 s, 344 s; U50,488H post-conditioning (post-CPP): −1,436 s, −109 s, 178 s; morphine pre-CPP: −542 s, −23 s, 413 s; morphine post-CPP: −299 s, 154 s, 819 s; C6 guano pre-CPP: −364 s, 43 s, 499 s; C6 guano post-CPP: −878 s, −42 s, 1,164 s; saline pre-CPP: −416 s, −93 s,

595 s; saline post-CPP: −411 s, 129 s, 438 s; vehicle pre-CPP: −377 s, 43 s, 573 s; vehicle post-CPP: −865 s, −75 s, 533 s. **f**, Comparison of C6 guano and U50,488H in an operant model of antinociception using CCI–CPP. A schematic representation of the CCI–CPP model protocol is presented in the Supplementary Methods. Points represent the difference in time spent on the drug-paired side. Data are mean ± s.e.m. U50,488H treatment in this model did not result in CPP or CPA, whereas C6 guano treatment resulted in CPP. Minimum, median and maximum values: U50,488H pre-CPP: −771 s, −52 s, 699 s; U50,488H post-CPP: −1,170 s, −67 s, 1,293 s; C6 guano pre-CPP: −767 s, 101 s, 438 s; C6 guano post-CPP: −566 s, 388 s, 1,397 s. **g**, Dose- and time-dependent anti-allodynic activity of C6 guano in a CCI model of neuropathic pain. Mechanical allodynia produced by sciatic nerve ligation was reduced between 20 and 100 min after treatment with 30 nmol or 100 nmol ICV C6 guano but not by 10 nmol ICV guano or vehicle. Data are mean ± s.e.m. **h**, Antinociception by C6 guano in the mouse acetic acid writhing test. Treatment with 100 nmol ICV C6 guano or 30 nmol ICV morphine showed significant antinociception compared with vehicle or saline. Data plotted are number of writhes counted over 15 min after administration of compound. Minimum, median and maximum values: vehicle: 7, 19.5, 37; morphine: 0, 0, 32; C6 guano: 0, 0.5, 17. **i**, Evaluation of C6 guano for antinociceptive effects in the formalin-induced inflammation assay in mouse. Treatment with 100 nmol ICV C6 guano or 30 nmol ICV morphine showed significant antinociception compared with vehicle or saline. Minimum, median and maximum values: vehicle: 34 s, 93.5 s, 151 s; morphine: 1 s, 7 s, 95 s; C6 guano: 0 s, 0 s, 78 s; saline: 29 s, 74 s, 150 s. *n* values and primary statistics for all panels are provided in Supplementary Table 2.

guano resulted in an increase, whereas morphine caused a decrease in breath rates with a 100 nmol ICV dose (Fig. 4d). Increased breath rates were also observed in μOR-knockout mice, suggesting that they were an off-target effect, independent of μOR. The mechanism underlying this increased breath rate remains unknown. An ideal μOR ligand would not have effects on respiration. However, in parallel testing to distinguish our ligand from a biased μOR agonist, we tested the respiratory effects of 7-hydroxymitragynine (7-OH). This ligand has μOR-dependent antinociceptive actions and retains hyperlocomotion as well as conditioned place preference (CPP) in mice[41]. Subcutaneous injection of 4.5 mg kg^{-1} 7-OH (15× the antinociceptive ED$_{50}$)

showed respiratory depression similar to that caused by morphine (30 mg kg^{-1} by subcutaneous injection; 15× the antinociceptive ED$_{50}$) (Supplementary Fig. 4d). The results are in agreement with results of oral administration of 7-OH[36]. Notably, other biased ligands[9,35,42] have also been shown to retain respiratory depression or provide marginal benefits over morphine.

Next, we tested C6 guano for reward or aversive behaviour in a CPP paradigm in mouse. ICV treatment with 100 nmol C6 guano did not induce CPP or conditioned place aversion (CPA); by contrast, morphine at an equianalgesic dose resulted in CPP and treatment with the κOR agonist U50,488H resulted in CPA (Fig. 4e).

For validation of antinociceptive effects beyond a thermal model, we tested C6 guano in models of operant pain, neuropathic pain (the chronic constriction injury (CCI) model), visceral pain (acetic acid writhing) and inflammatory pain (the formalin assay). In an operant pain model that subjected mice exposed to CCI to place conditioning in the CPP assay[43], 100 nmol ICV U50,488H did not result in CPP or CPA, consistent with previous reports with systemic U50,488H[43]. By contrast, place conditioning with 100 nmol ICV C6 guano resulted in CPP, suggesting that it can blunt the aversive emotions associated with neuropathic pain (Fig. 4f). Extending this finding, treatment with C6 guano dose-dependently ameliorated mechanical allodynia displayed by mice exposed to CCI, with near-maximal efficacy and a longer time course against this model of neuropathic pain than observed with the thermal pain assays (Fig. 4g). Moreover, mice given 100 nmol ICV C6 guano demonstrated antinociception equivalent to that of 30 nmol ICV morphine in the writhing assay (Fig. 4h) and in the formalin assay (Fig. 4i).

Overall, our in vivo tests suggest that C6 guano produced antinociception against diverse pain types but did not show typical μOR-mediated adverse effects.

Discussion

Functional selectivity has been proposed as a path to more precise pharmacology, to enable the separation of therapeutic efficacy from side effects in several receptor families, including opioid[9], angiotensin[44], adrenergic, dopamine, serotonin and other receptors[45].

Here we used bitopic extensions of the fentanyl scaffold to design functionally selective ligands for μOR, a typical family A GPCR. Our results demonstrate that guanidine extension of the ligand (C5 guano and C6 guano) can engage the conserved $D^{2.50}$ side chain of the sodium site, either directly via a salt bridge or indirectly via water. Binding of the guanidine moiety in the sodium pocket can markedly reduce or even abolish β-arrestin-1 and β-arrestin-2 recruitment compared with the parent ligand fentanyl, while maintaining G-protein partial agonism in less amplified BRET-based assays with limited receptor reserve. We also observed a distinct G_i, G_o and G_z activity profile of C6 guano, which engages the $D^{2.50}$ residue directly, compared with C5 guano, which engages the $D^{2.50}$ site more indirectly via water. Notably, there was a selective decrease in efficacy in G_z with C6 guano compared with the other tested compounds (Fig. 3j), and a narrower spread of Gα-subtype potency compared with C5 guano (Fig. 3e) and partial agonists such as pentazocine and buprenorphine, full agonists such as DAMGO, fentanyl, carfentanil and other G-protein-biased μOR partial agonists, morphinans and opioid peptides. It should be noted that replacement of the *n*-aniline ring with small aliphatic groups[46] has been shown to result in abolished β-arrestin recruitment; however, this also reduces G-protein efficacy (by 60–70%) and potency (80–1,000-fold) compared with the parent fentanyl. By contrast, our C5 and C6 bitopic ligands maintained fentanyl-like high affinity as well as cAMP potency and efficacy. It is possible that the guanidine functional group that engages $D^{2.50}$ can differentially affect the conformation and/or dynamics of the helices involved in receptor activation. Thus, the bulky guanidine fitting the sodium pocket may be able to block the well-characterized inward shift of transmembrane helix (TM)7 previously associated with arrestin signalling, while simultaneously allowing the outward shift of TM6 that opens the G-protein-binding site[44,47].

It is currently believed that the analgesia, respiratory depression and physical dependence mediated by the μOR are all dependent on G-protein pathways[39,48] and not on β−arrestin-2 as previously proposed[34]. μOR is coupled to six distinct G-protein isoforms, and the possibility that functional selectivity for some of these isoforms can be used to separate analgesia from adverse effects remains largely unexplored. As our results suggest, the bitopic design that extends orthosteric ligands such as fentanyl into the conserved sodium pocket may provide a platform for the development of such Gα-subtype-biased ligands.

Several reports have suggested that G_i, G_o and G_z subtypes may have different roles in the physiological responses to μOR agonists. For example, in studies using antisense oligonucleotides to reduce the expression of individual G_i, G_o and G_z subtypes, DAMGO elicits activation of G_{i2}, G_z and G_q, whereas endomorphin-1 involves the activation of G_{i1}, G_{i3} and G_z to produce supraspinal analgesia, suggesting that different Gα subunits mediate analgesia in response to these two μOR ligands[49–51]. The role of G_z in opioid-induced antinociception and adverse effects is less well understood. G_z-knockout animals have shown little or no change in supraspinal analgesic effects of morphine[52], yet they exhibit a marked increase in analgesic tolerance[53,54] and a decrease in lethality[54]. Studies in G_o-knockout mice suggest that supraspinal analgesia is primarily mediated by one or both G_o subtypes. Homozygous G_o-knockout mice are not viable; however, supraspinal analgesia in heterozygous G_o-knockout mice is diminished, even though spinal analgesic actions of morphine remain intact[55]. Supraspinal analgesia of morphine is intact in mice deficient in G_{i2} or G_{i3}[55]. To our knowledge, there are no reported studies in G_{i1}-knockout mice, possibly owing to lethality in homozygotes. Of note, studies using ICV injection of antisense oligonucleotides to suppress G_i, G_o and G_z subtype expression have shown that all G_i, G_o and G_z antisense oligonucleotides suppress the analgesia produced by heroin, suggesting all subtypes contribute to analgesia[51]. Thus, the physiological roles of Gα-subtypes are incompletely understood and compounds with distinct G_i, G_o and G_z signalling profiles may enable us to assess the roles of specific subtypes in the therapeutic and adverse effects of μOR agonists. To this end, C6 guano could be a valuable tool and provide a platform to probe the pharmacology of G_z coupled to μOR. When evaluated in animal models, C6 guano exhibited supraspinal analgesia mediated by μOR in a range of rodent pain models while showing attenuated abuse potential and aversion. It is possible that reduced G_z efficacy and/or lower efficacy at all Gα-subtypes and/or other opioid targets is responsible for this distinct in vivo profile of C6 guano. Unexpectedly, C6 guano caused an increase in respiration rate via an unknown mechanism independent of μOR activation.

In summary, we have demonstrated that functional selectivity of the fentanyl scaffold can be controlled by bitopic extension of the ligand into the μOR sodium pocket. This is consistent with reports in which the highly conserved allosteric pocket has a central role in the functional mechanism of family A GPCRs[2,56]. Previously, major changes in receptor basal activity[18,57–59], efficacy[2,56–59] and functional selectivity switches[2,18,56] that can bias GPCR signalling towards either G_i-protein or β-arrestin-2 pathways were observed for point mutations in the Na$^+$-binding pocket for a variety of family A receptors. Our study reveals that such functional modulation can be achieved with bitopic ligands specifically designed to interact with the sodium pocket. This design was validated by cryo-EM structural studies and detailed analysis of the distinct signalling profiles at μOR of the bitopic ligands. Moreover, our bitopic ligands were able to modulate functional selectivity both between G protein and β-arrestin-2, as well as between G-protein subtypes, resulting in a new class of antinociceptive agents with attenuated adverse opioid effects. Exceptional conservation of the sodium pocket across GPCR structures[2] suggests that such bitopic ligands could be designed for many other family A GPCRs. Although the functional effects would differ between different receptors, the rational design of such bitopic ligands could provide highly versatile pharmacological probes for exploring the multidimensional signalling landscape of GPCRs.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions

Article

and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-05588-y.

1. DeWeerdt, S. Tracing the US opioid crisis to its roots. *Nature* **573**, S10–S12 (2019).
2. Zarzycka, B., Zaidi, S. A., Roth, B. L. & Katritch, V. Harnessing ion-binding sites for GPCR pharmacology. *Pharmacol. Rev.* **71**, 571–595 (2019).
3. Huang, W. et al. Structural insights into μ-opioid receptor activation. *Nature* **524**, 315–321 (2015).
4. Hilger, D., Masureel, M. & Kobilka, B. K. Structure and dynamics of GPCR signaling complexes. *Nat. Struct. Mol. Biol.* **25**, 4–12 (2018).
5. Pasternak, G. W. & Pan, Y.-X. Mu opioids and their receptors: evolution of a concept. *Pharmacol. Rev.* **65**, 1257–317 (2013).
6. Varga, B. R., Streicher, J. M. & Majumdar, S. Strategies towards safer opioid analgesics—a review of old and upcoming targets. *Br. J. Pharmacol.* https://doi.org/10.1111/bph.15760 (2021).
7. Manglik, A. et al. Structure-based discovery of opioid analgesics with reduced side effects. *Nature* **537**, 185–190 (2016).
8. DeWire, S. M. et al. A G protein-biased ligand at the μ-opioid receptor is potently analgesic with reduced gastrointestinal and respiratory dysfunction compared with morphines. *J. Pharmacol. Exp. Ther.* **344**, 708–717 (2013).
9. Faouzi, A., Varga, B. R. & Majumdar, S. Biased opioid ligands. *Molecules* **25**, 4257 (2020).
10. Uprety, R. et al. Controlling opioid receptor functional selectivity by targeting distinct subpockets of the orthosteric site. *eLife* **10**, e56519 (2021).
11. Schmid, C. L. et al. Bias factor and therapeutic window correlate to predict safer opioid analgesics. *Cell* **171**, 1165–1175.e13 (2017).
12. Eans, S. O. et al. Parallel synthesis of hexahydrodiimidazodiazepines heterocyclic peptidomimetics and their in vitro and in vivo activities at μ (MOR), δ (DOR), and κ (KOR) opioid receptors. *J. Med. Chem.* **58**, 4905–4917 (2015).
13. Majumdar, S. et al. Truncated G protein-coupled mu opioid receptor MOR-1 splice variants are targets for highly potent opioid analgesics lacking side effects. *Proc. Natl Acad. Sci. USA* **108**, 19778–19783 (2011).
14. Kiguchi, N. et al. BU10038 as a safe opioid analgesic with fewer side-effects after systemic and intrathecal administration in primates. *Br. J. Anaesth.* **122**, e146–e156 (2019).
15. Váradi, A. et al. Mitragynine/corynantheidine pseudoindoxyls as opioid analgesics with mu agonism and delta antagonism, which do not recruit β-arrestin-2. *J. Med. Chem.* **59**, 8381–8397 (2016).
16. Massaly, N., Temp, J., Machelska, H. & Stein, C. Uncovering the analgesic effects of a pH-dependent mu-opioid receptor agonist using a model of nonevoked ongoing pain. *Pain* **161**, 2798–2804 (2020).
17. Kandasamy, R. et al. Positive allosteric modulation of the mu-opioid receptor produces analgesia with reduced side effects. *Proc. Natl Acad. Sci. USA* **118**, e2000017118 (2021).
18. Fenalti, G. et al. Molecular control of δ-opioid receptor signalling. *Nature* **506**, 191–196 (2014).
19. Pert, C. B., Pasternak, G. & Snyder, S. H. Opiate agonists and antagonists discriminated by receptor binding in brain. *Science* **182**, 1359–1361 (1973).
20. Hu, X. et al. Kinetic and thermodynamic insights into sodium ion translocation through the μ-opioid receptor from molecular dynamics and machine learning analysis. *PLoS Comput. Biol.* **15**, e1006689 (2019).
21. Shang, Y. et al. Mechanistic insights into the allosteric modulation of opioid receptors by sodium ions. *Biochemistry* **53**, 5140–5149 (2014).
22. Liu, W. et al. Structural basis for allosteric regulation of GPCRS by sodium ions. *Science* **337**, 232–236 (2012).
23. Selvam, B., Shamsi, Z. & Shukla, D. Universality of the sodium ion binding mechanism in class A G-protein-coupled receptors. *Angew. Chem. Int. Ed. Engl.* **57**, 3048–3053 (2018).
24. Manglik, A. et al. Crystal structure of the μ-opioid receptor bound to a morphinan antagonist. *Nature* **485**, 321–326 (2012).
25. Chakraborty, S. et al. A novel mitragynine analog with low-efficacy mu opioid receptor agonism displays antinociception with attenuated adverse effects. *J. Med. Chem.* **64**, 13873–13892 (2021).
26. Overdose Death Rates. *National Institute on Drug Abuse* https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates (accessed 23 September 2019).
27. Lipiński, P. F. J., Jarończyk, M., Dobrowolski, J. C. & Sadlej, J. Molecular dynamics of fentanyl bound to μ-opioid receptor. *J. Mol. Model.* **25**, 144 (2019).
28. Subramanian, G., Paterlini, M. G., Portoghese, P. S. & Ferguson, D. M. Molecular docking reveals a novel binding site model for fentanyl at the mu-opioid receptor. *J. Med. Chem.* **43**, 381–391 (2000).
29. Qu, Q. et al. Structural insights into distinct signaling profiles of the μOR activated by diverse agonists. *Nat. Chem. Biol.* https://doi.org/10.1038/s41589-022-01208-y (2022).
30. Ballasteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Meth. Neurosci.* **25**, 366–428 (1995).
31. Dardonville, C. et al. Synthesis and pharmacological studies of new hybrid derivatives of fentanyl active at the μ-opioid receptor and I₂-imidazoline binding sites. *Bioorg. Med. Chem.* **14**, 6570–6580 (2006).
32. Gallivan, J. P. & Dougherty, D. A. Cation–π interactions in structural biology. *Proc. Natl Acad. Sci. USA* **96**, 9459–9464 (1999).
33. Olsen, R. H. J. et al. TRUPATH, an open-source biosensor platform for interrogating the GPCR transducerome. *Nat. Chem. Biol.* **16**, 841–849 (2020).
34. Raehal, K. M., Walker, J. K. L. & Bohn, L. M. Morphine side effects in β-arrestin 2 knockout mice. *J. Pharmacol. Exp. Ther.* **314**, 1195–1201 (2005).
35. Gillis, A. et al. Low intrinsic efficacy for g protein activation can explain the improved side effect profiles of new opioid agonists. *Sci. Signal.* **13**, 31 (2020).
36. Hill, R., Kruegel, A. C., Javitch, J. A., Lane, J. R. & Canals, M. The respiratory depressant effects of mitragynine are limited by its conversion to 7-OH mitragynine. *Br. J. Pharmacol.* **179**, 3875–3885 (2022).
37. He, L. et al. Pharmacological and genetic manipulations at the μ-opioid receptor reveal arrestin-3 engagement limits analgesic tolerance and does not exacerbate respiratory depression in mice. *Neuropsychopharmacology* **46**, 2241–2249 (2021).
38. Anesthetic and Analgesic Drug Products Advisory Committee. *Oliceridine* FDA Advisory Committee Briefing Document (FDA, 2018).
39. Kliewer, A. et al. Phosphorylation-deficient G-protein-biased μ-opioid receptors improve analgesia and diminish tolerance but worsen opioid side effects. *Nat. Commun.* **10**, 367 (2019).
40. Besnard, J. et al. Automated design of ligands to polypharmacological profiles. *Nature* **492**, 215–220 (2012).
41. Chakraborty, S. et al. Oxidative metabolism as a modulator of kratom's biological actions. *J. Med. Chem.* **64**, 16553–16572 (2021).
42. Hill, R. et al. The novel μ-opioid receptor agonist PZM21 depresses respiration and induces tolerance to antinociception. *Br. J. Pharmacol.* **175**, 2653 (2018).
43. Wilson, L. L. et al. Characterization of CM-398, a novel selective sigma-2 receptor ligand, as a potential therapeutic for neuropathic pain. *Molecules* **27**, 3617 (2022).
44. Wingler, L. et al. Angiotensin analogs with divergent bias stabilize distinct receptor conformations. *Cell* **176**, 468–478.e11 (2019).
45. Wootten, D., Christopoulos, A., Marti-Solano, M., Babu, M. M. & Sexton, P. M. Mechanisms of signalling and biased agonism in G protein-coupled receptors. *Nat. Rev. Mol. Cell Biol.* **19**, 638–653 (2018).
46. de Waal, P. W. et al. Molecular mechanisms of fentanyl mediated β-arrestin biased signaling. *PLoS Comput. Biol.* **16**, e1007394 (2020).
47. Liu, J.J, Horst, R, Katritch, V, Stevens, R.C., & Wuthrich, K. Biased signaling pathways in β2-adrenergic receptor characterized by ¹⁹F-NMR. *Science* **335**, 1106–1110 (2012).
48. Kliewer, A. et al. Morphine-induced respiratory depression is independent of β-arrestin2 signalling. *Br. J. Pharmacol.* **177**, 2923–2931 (2020).
49. Standifer, K. M., Rossi, G. C. & Pasternak, G. W. Differential blockade of opioid analgesia by antisense oligodeoxynucleotides directed against various G protein alpha subunits. *Mol. Pharmacol.* **50**, 293–298 (1996).
50. Sánchez-Blázquez, P., Rodríguez-Díaz, M., DeAntonio, I. & Garzón, J. Endomorphin-1 and endomorphin-2 show differences in their activation of μ opioid receptor-regulated G proteins in supraspinal antinociception in mice. *J. Pharmacol. Exp. Ther.* **291**, 12–18 (1999).
51. Sánchez-Blázquez, P., Gómez-Serranillos, P. & Garzón, J. Agonists determine the pattern of G-protein activation in μ-opioid receptor-mediated supraspinal analgesia. *Brain Res. Bull.* **54**, 229–235 (2001).
52. Yang, J. et al. Loss of signaling through the G protein, Gz, results in abnormal platelet activation and altered responses to psychoactive drugs. *Proc. Natl Acad. Sci. USA* **97**, 9984–9989 (2000).
53. Hendry, I. A. et al. Hypertolerance to morphine in G(zα)-deficient mice. *Brain Res.* **870**, 10–19 (2000).
54. Leck, K. J. et al. Deletion of guanine nucleotide binding protein αz subunit in mice induces a gene dose dependent tolerance to morphine. *Neuropharmacology* **46**, 836–846 (2004).
55. Lamberts, J. T., Jutkiewicz, E. M., Mortensen, R. M. & Traynor, J. R. Mu-opioid receptor coupling to Gαz plays an important role in opioid antinociception. *Neuropsychopharmacology* **36**, 2041–2053 (2011).
56. Katritch, V. et al. Allosteric sodium in class A GPCR signaling. *Trends Biochem. Sci* **39**, 233–244 (2014).
57. Schöppe, J. et al. Crystal structures of the human neurokinin 1 receptor in complex with clinically used antagonists. *Nat. Commun.* **10**, 17 (2019).
58. Massink, A. et al. Sodium ion binding pocket mutations and adenosine A₂ₐ receptor function. *Mol. Pharmacol.* **87**, 305–313 (2015).
59. Capaldi, S. et al. Allosteric sodium binding cavity in GPR3: a novel player in modulation of aβ production. *Sci. Rep.* **8**, 11102 (2018).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The cryo-EM maps and corresponding coordinates have been deposited in the Electron Microscopy Data Bank (EMDB) under accession codes EMD-26314 (C5 guano–μOR–G_i–scFv16) and EMD-26313 (C6 guano–μOR–G_i) and the Protein Data Bank (PDB) under accession codes 7U2L (C5 guano–μOR–G_i–scFv16) and 7U2K (C6 guano–μOR–G_i). The authors declare that all the data supporting the findings of this study are available within the article, extended data and supplementary information files. All compounds can be made available on reasonable requests from the authors. Source data are provided with this paper.

Author contributions S.M., B.K.K., V.K. and G.S. conceived the study. A.F. and B.R.V. synthesized the compounds, aided in their characterization under the supervision of S.M. H.W. prepared the μOR–G_i complex, obtained and processed cryo-EM data, and refined the structure from cryo-EM density maps under the supervision of B.K.K. and G.S. S.A.Z. performed the docking, ligand design, and molecular dynamics simulations under the supervision of V.K. Q.Q. and M.J.R obtained and processed cryo-EM data under the supervision of G.S. S.T.S. and J.F.D. performed the profiling studies under the supervision of B.L.R. A.E.D. performed the profiling studies under the supervision of T.C. K.A. carried out TRUPATH and pharmacokinetics assays under the supervision of S.M. T.Z. carried out binding assays under the supervision of Y.X.P. S.L. and J.X. performed the antinociception assay with ICV administration under the supervision of Y.X.P. C.R. carried out mouse brain stability assays under the supervision of M.D.C. S.O.E. and M.K.M. carried out behavioural assays under the supervision of J.P.M. and R.A.H., respectively. A.F., H.W., S.A.Z., J.F.D, J.P.M., V.K., G.S., B.K.K. and S.M. wrote the paper with contributions from the other authors.

Competing interests S.M. and Y.X.P. are founders of Sparian Biosciences. B.K.K. is a founder of and consultant for ConfometRx. G.S. is a cofounder of Deep Apple. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-05588-y.

Correspondence and requests for materials should be addressed to Georgios Skiniotis, Vsevolod Katritch, Brian K. Kobilka or Susruta Majumdar.

Peer review information *Nature* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints.

Article
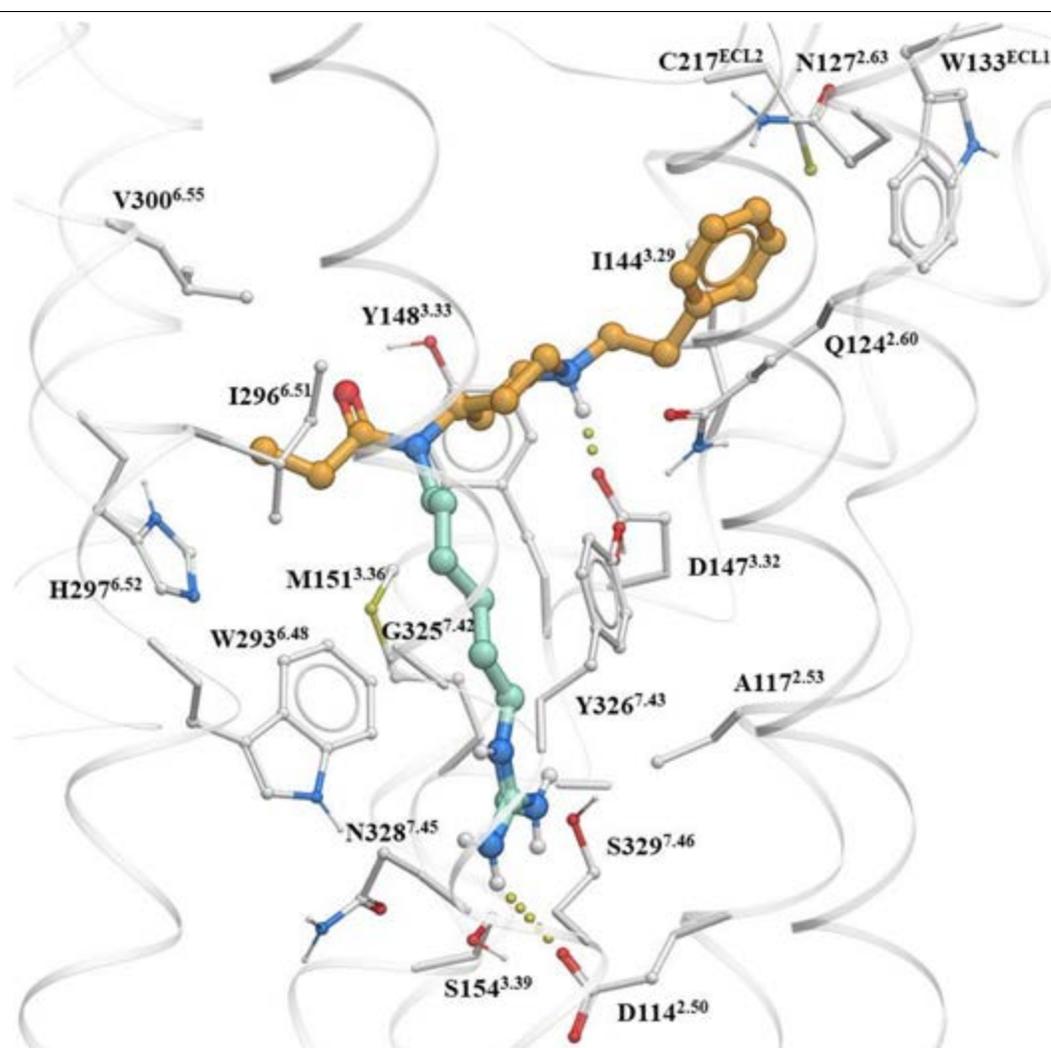
Extended Data Fig. 1 | Docking of a fentanyl based bitopic targeting the Na⁺ binding site. Molecular docking of a fentanyl based bitopic ligand shows that the functional head group can target the Na⁺ pocket.

A 3470 micrographs → motion correction / ctf correction / autopicking → 2,037,265 particles → 2D classification → 784,061 particles

3D classification

CtfRefine / Particle polishing / 3D classification without image alignment → 373,936 particles

3D refinement / Postprocess → 164,647 particles → 3.2Å

B 3019 micrographs → motion correction / ctf correction / autopicking → 994,998 particles → 2D classification → 489,212 particles

3D classification

CtfRefine / Particle polishing / 3D classification without image alignment → 436,329 particles

3D refinement / Postprocess → 232,146 particles

3.3Å → Substract scFv and micelle / 3D refinement / Postprocess → 3.3Å

Extended Data Fig. 2 | Cryo-EM data processing work-flows. Representative micrographs, 2D classes, 3D classes and data processing procedures for (A) C5-guano and (B) C6-guano bound µOR–G$_i$-scFv16 complex.

Article

A

B
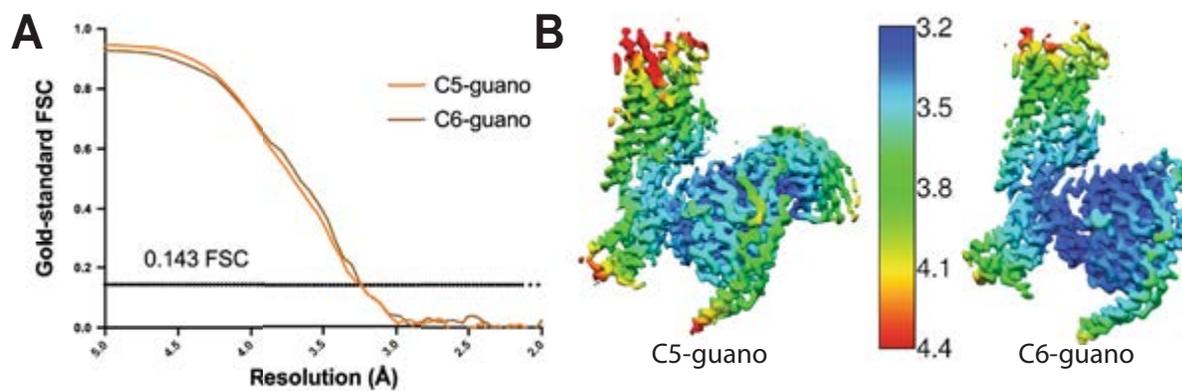
C5-guano   C6-guano

Extended Data Fig. 3 | Global and local resolutions for cryo-EM maps. (A) Gold-standard FSC curves for C5-guano and C6-guano bound μOR–G_i structures. Overall resolution is 3.2 Å for C5-guano bound μOR–G_i-scFv16 and 3.3 Å for C6-guano bound μOR–G_i using the gold Standard FSC = 0.143 criterion. (B) Local resolution map of C5 guano and C6 guano bound μOR–G_i structures. (C) Data collection, refinement, and model statistic of two structures. Extended Data Table 2. Cryo-EM data collection, refinement and validation statistics.
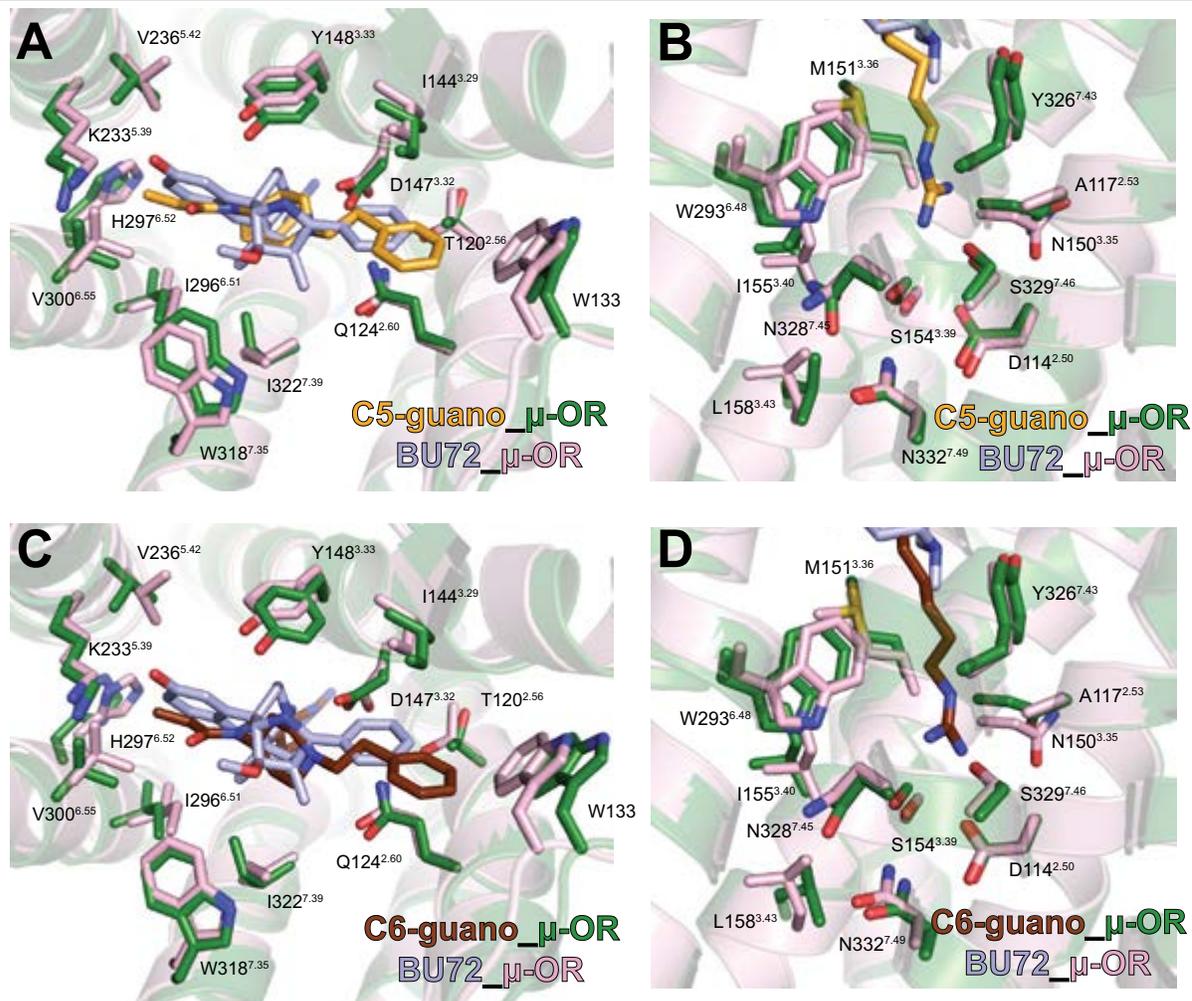
Extended Data Fig. 4 | Comparison of bitopic structures to BU72 structure. A, C, Side chains of μOR orthosteric pocket residues are shown for the C5-guano (A) and C6-guano (C) bound μOR–G$_i$ complex (green) in comparison with the BU72 bound μOR (PDB code 5C1M; pink). The orthosteric pocket residues of μOR in complex with bitopic ligands and BU72 show nearly identical conformations. B, D, Side chains of μOR site-2 and Na$^+$ site residues are shown for the C5 guano (B) and C6 guano (D) bound μOR–G$_i$ complex (green) in comparison with the BU72 bound μOR (PDB code 5C1M; pink). The site-2 and Na$^+$ site residues of μOR in complex with bitopic ligands and BU72 show nearly identical conformations.

Article

F

Types of Guanidine interactions with D2.50				
1 water-mediated bridge, % of frames	2 water-mediated bridges, % of frames	Direct interaction with 1 supplementary water bridge, % of frames	Direct interaction with 2 supplementary water bridges, % of frames	Direct interaction with no water bridges, % of frames
41	1	5.2	0.53	8.2
7.2	6.9	44	0.15	40
64	5.6	8.8	0.30	5.4
19	2.4	4.6	0.12	67
1.4	0.19	2.8	12	82
4.5	2.3	14	0.17	79
2.6	0.62	0.60	1.47	94
27	3.5	25	0.48	38
47	20	5.3	0.00	12
54	9.9	12	0.05	12
27	**6.3**	**12**	**1.6**	**44**

G

Types of Guanidine interactions with D2.50				
1 water-mediated bridge, % of frames	2 water-mediated bridges, % of frames	Direct interaction with 1 supplementary water bridge, % of frames	Direct interaction with 2 supplementary water bridges, % of frames	Direct interaction with no water bridges, % of frames
2.35	1.1	44	0.23	52
0.38	0.01	61	0.03	30
0.89	0.03	16	0.03	83
35.2	11	37	0.79	8.9
3.5	0.49	62	0.29	33
8.5	**2.6**	**44**	**0.27**	**41**

Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Analysis of dynamics of direct and water mediated interactions of bitopic ligands. A) Overlay of three examples of C5 guano conformations bound to active state MOR (pink cartoon/sticks) during MD simulations (B) Detailed view of the interactions between guano moiety of C5 guano (orange sticks) and D114[2.50] mediated by two water molecule (C) Direct salt bridge interactions between C5 guano (light green sticks) and D116[2.50] supplemented by an additional water-mediated hydrogen bond. (D) direct salt bridge interactions between C5 guano (cyan sticks) and D114[2.50] (E) Probability densities of distances between guano nitrogen atoms and D114[2.50] carboxylate oxygens. Each chart plots probability density for frames with two bridging waters (orange), one bridging water (green), and no bridging waters (cyan). (F) Categorization and relative proportion of D2.50 and D3.32 mediated interactions in 10 independent C5 guano-µOR MD trajectories for 1 µs each.

Among the cumulative frames from the 10 µs MD runs, close to 1/3rd of the frames-maintained guano-D2.50 interactions exclusively through water-mediated hydrogen bonds, while ~57% frames formed direct salt- bridges with or without supplementary water mediated interactions. Therefore, close to 90% of the frames maintained D2.50-guano interactions. The piperidine-D3.32 interactions were observed to be even more stable, with over 96% of the frames indicating direct salt bridge or water-mediated hydrogen bonds. (G) Categorization and relative proportion of D2.50 and D3.32 mediated interactions in 5 independent C6-µOR trajectories for 1 µs each. Overall, the number of direct interactions with D2.50 increased from 57% to 85% (compared to C5), perhaps resulting from the increase in linker length by a carbon atom that decreases the overall distances to D2.50 residue.

Article
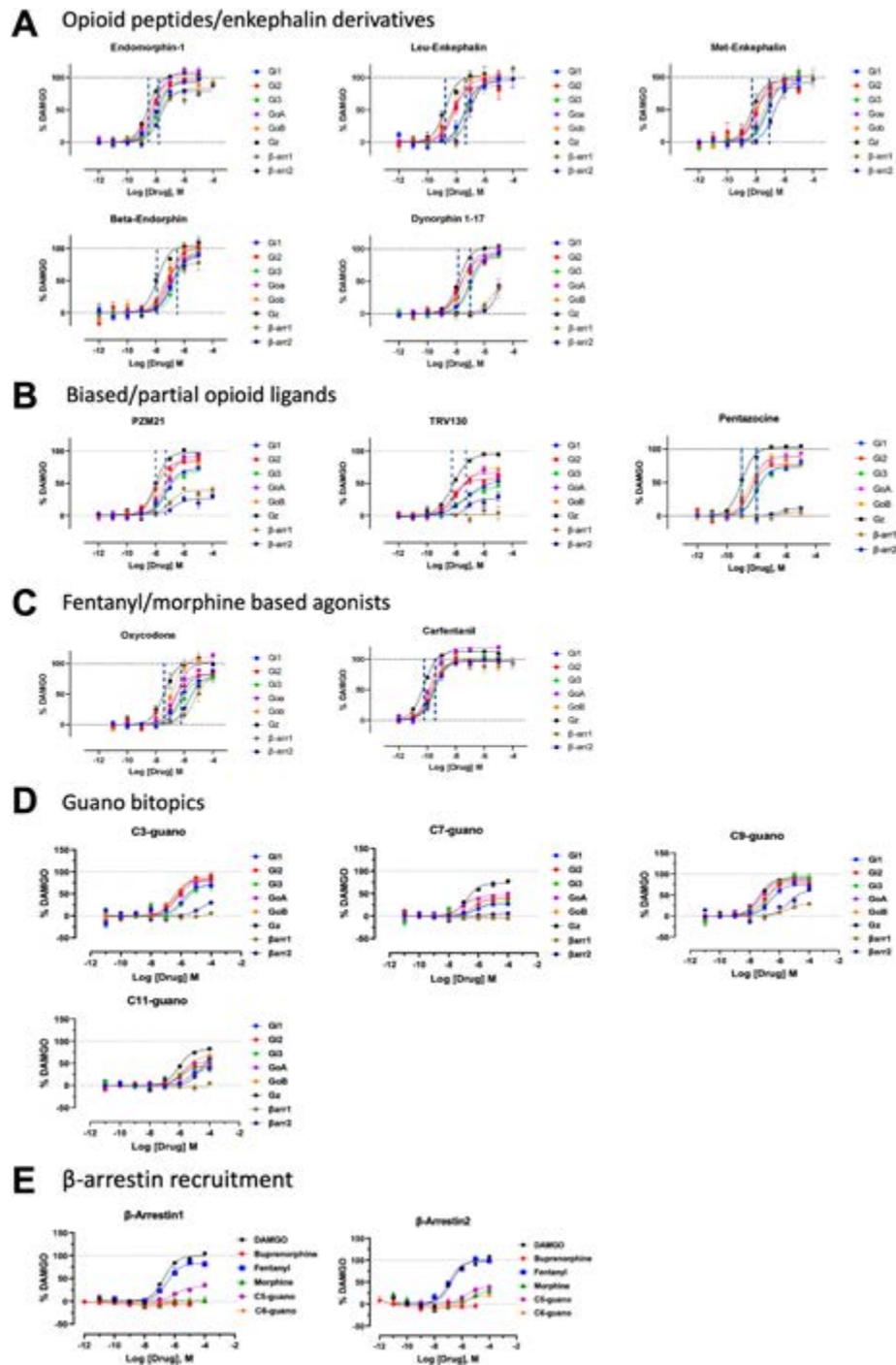
Extended Data Fig. 6 | Profiling of chemically and pharmacologically distinct μOR agonists using TRUPATH, arrestin signaling. A) Peptides: Endomorphin-1, Leu-enkephalin, Met-enkephalin, Beta-endorphin and Dynorphin A (1-17). Dynorphin A (1-17) showed reduced arrestin recruitment while other peptides retained robust arrestin recruitment among peptides tested. B) Opioid biased agonists and partials: PZM21, TRV130, Gα-subtype selectivity and arrestin recruitment on μOR. PZM21, 7-OH and TRV130 showed <50% efficacy for arrestin1/2. Highest efficacy for all three biased agonists was seen at the G_z-subtype. μOR partial agonist pentazocine was a full agonist at the G_z subtype. C) Oxycodone and Carfentanil, Gα-subtype selectivity and arrestin recruitment on μOR. Carfentanil showed near maximal efficacy at all Gα-subtypes and arrestin1/2. Oxycodone was a full agonist at G_z and showed >50% efficacy at β-arrestin2. D) Fentanyl guano bitopics show differential G-protein and arrestin efficacy with increased chain length.

Extended Data Table 1 | Summary of binding affinities, cAMP and arrestin recruitment values of fentanyl amino and guano bitopics on µOR

Ligands	Binding [a]	cAMP[b]			β-arrestin2[b]		
	$K_i \pm$ SEM (nM)	EC_{50} (nM)	$pEC_{50} \pm$ SEM	$E_{max} \pm$ SEM (%)	EC_{50} (nM)	$pEC_{50} \pm$ SEM	$E_{max} \pm$ SEM (%)
C3 amino	> 1000	12300	4.9 ± 0.2	142 ± 20	n.d.	n.d.	8.2 ± 1.1
C3 guano	77 ± 3.0	164	6.8 ± 0.1	107 ± 2.3	n.d.	n.d.	6.9 ± 2.4
C5 amino	1179 ± 13	469	6.3 ± 0.2	90 ± 5.9	n.d.	n.d.	n.d.
C5 guano	4.6 ± 0.7	0.8	9.1 ± 0.1	104 ± 3.1	4710	5.3 ± 0.1	54 ± 3.2
C6 amino	1033 ± 15	844	6.1 ± 0.1	126 ± 4.7	n.d.	n.d.	20 ± 2.3
C6 guano	4.1 ± 0.3	4.7	8.3 ± 0.1	137 ± 2.1	n.d.	n.d.	19 ± 1.4
C7 amino	369 ± 7.0	436	6.4 ± 0.1	109 ± 3.9	n.d.	n.d.	n.d.
C7 guano	41 ± 8.0	62	7.2 ± 0.1	109 ± 2.4	n.d.	n.d.	13 ± 1.4
C9 amino	281 ± 2.2	360	6.4 ± 0.1	106 ± 4.6	66720	4.2 ± 0.2	56 ± 9.4
C9 guano	58 ± 8.0	21.3	7.7 ± 0.1	128 ± 3.2	792	5.1 ± 0.1	120 ± 6.6
C11 amino	589 ± 13	293.2	6.5 ± 0.1	105 ± 3.6	n.d.	n.d.	n.d
C11 guano	165 ± 4	1020	6.0 ± 0.1	118 ± 5.2	n.d.	n.d.	n.d.
Fentanyl	9.1 ± 3.1	0.90	9.0 ± 0.1	86 ± 1.8	66.2	7.2 ± 0.1	119 ± 2.7
DAMGO	n.d.	0.40	9.4 ± 0.1	99 ± 4	723	6.1 ± 0.1	100 ± 2.8

[a]K_i binding competition studies were performed with the indicated compound against ^{125}I-IBNtxA (0.1 nM) in membranes from CHO cells stably expressing the µOR cloned mouse opioid receptor. Results are presented as $K_i \pm$ SEM (nM) from three independent experiments performed in triplicate. [b]Potency and efficacy data were obtained using agonist induced inhibition measured by cyclic AMP (cAMP) and Tango assay for β-arrestin 2 recruitment, respectively, from three independent experiments performed in triplicate. Efficacy is represented as EC_{50} (nM) and percent maximal stimulation (E_{max}) relative to standard agonist DAMGO. "n.d." Denotes not determined. Efficacy<20%.

Article

Extended Data Table 2 | Cryo-EM data collection, refinement and validation statistics

	C5-guano-μOR-Gi-scFv16 (EMDB-26314) (PDB 7U2L)	C6-guano-μOR-Gi (EMDB-26313) (PDB 7U2K)
Data collection and processing		
Magnification	130,000	130,000
Voltage (kV)	300	300
Electron exposure (e–/Å^2)	67	68.5
Defocus range (μm)	0.7-2	0.7-2
Pixel size (Å)	1.06	0.8676
Symmetry imposed	C1	C1
Initial particle images (no.)	2,037,265	994,998
Final particle images (no.)	164,647	232,146
Map resolution (Å)	3.2	3.3
FSC threshold	0.143	0.143
Map resolution range (Å)	3.2-4.9	3.2-4.6
Refinement		
Initial model used (PDB code)	6DDE	6DDE
Model resolution (Å)	3.2	3.2
FSC threshold	0.143	0.143
Model resolution range (Å)	2.9-3.4	2.8-3.4
Map sharpening B factor (Å^2)	-161.76	-103.46
Model composition		
Non-hydrogen atoms	8739	6759
Protein residues	1122	889
Ligands	1	1
B factors (Å^2)		
Protein	60.76	92.70
Ligand	78.28	125.85
R.m.s. deviations		
Bond lengths (Å)	0.004	0.009
Bond angles (°)	0.566	0.768
Validation		
MolProbity score	1.72	1.78
Clashscore	8.03	7.07
Poor rotamers (%)	0	0
Ramachandran plot		
Favored (%)	95.93	94.18
Allowed (%)	4.07	5.82
Disallowed (%)	0	0

Statistics of the cryo-EM maps and coordinates.

Extended Data Table 3 | Potency table for drugs profiled in Extended Data Fig. 6

Potency

Compounds	G_{i1} EC$_{50}$ nM (pEC$_{50}$ ± SEM)	G_{i2} EC$_{50}$ nM (pEC$_{50}$ ± SEM)	G_{i3} EC$_{50}$ nM (pEC$_{50}$ ± SEM)	G_{oa} EC$_{50}$ nM (pEC$_{50}$ ± SEM)	G_{ob} EC$_{50}$ nM (pEC$_{50}$ ± SEM)	G_z EC$_{50}$ nM (pEC$_{50}$ ± SEM)	β-arr1 EC$_{50}$ nM (pEC$_{50}$ ± SEM)	β-arr2 EC$_{50}$ nM (pEC$_{50}$ ± SEM)
Opioid Peptides								
DAMGO	18 (7.7 ± 0.1)	5 (8.2±0.1)	15 (7.8±0.1)	5 (8.3±0.1)	2.5 (8.6±0.1)	5.6 (9.2±0.1)	189 (6.7±0.1)	171 (6.8±0.1)
Endomorphin-1	9.9 (8.0 ± 0.1)	4.3 (8.4 ± 0.1)	16 (7.8 ± 0.1)	8.0 (8.1 ± 0.1)	6.0 (8.2 ± 0.1)	3.0 (8.5 ± 0.1)	11 (7.9 ± 0.1)	21 (7.7 ± 0.1)
Leu-Enkephalin	39 (7.4 ± 0.1)	5.1 (8.3 ± 0.2)	47 (7.3 ± 0.1)	8.6 (8.1 ± 0.1)	6.7 (8.2 ± 0.1)	1.8 (8.8 ± 0.1)	98 (7.0 ± 0.2)	90 (7.1 ± 0.2)
Met-Enkephalin	54 (7.3 ± 0.1)	10 (8.00 ± 0.3)	79 (7.1 ± 0.1)	12 (7.9 ± 0.1)	8.6 (8.1 ± 0.1)	5.3 (8.3 ± 0.1)	39 (7.4 ± 0.2)	193 (6.7 ± 0.2)
Beta-Enkephalin	128 (6.9 ± 0.1)	93 (7.0 ± 0.2)	293 (6.5 ± 0.1)	62 (7.2 ± 0.1)	55 (7.3 ± 0.2)	13 (7.9 ± 0.1)	28 (7.6 ± 0.2)	120 (6.9 ± 0.1)
Dynorphin 1-17	88 (7.1 ± 0.1)	18 (7.7 ± 0.1)	115 (6.7 ± 0.1)	25 (7.6 ± 0.1)	43 (7.4 ± 0.1)	167 (7.8 ± 0.1)	nd	nd
Biased Opiates								
PZM21	53 (7.3 ± 0.07)	10 (8.0 ± 0.1)	55 (7.3 ± 0.1)	52 (7.3 ± 0.1)	28 (7.6 ± 0.1)	9.8 (8.0 ± 0.1)	104 (7.0 ± 0.2)	199 (6.7 ± 0.4)
7-OH	111 (6.9 ± 0.3)	76 (7.1 ± 0.2)	nd	39 (7.4 ± 0.2)	55 (7.3 ± 0.2)	8.6 (8.1 ± 0.1)	nd	nd
TRV130	51 (7.3 ± 0.1)	5.9 (8.2 ± 0.2)	26 (7.6 ± 0.2)	12 (7.9 ± 0.2)	18 (7.7 ± 0.1)	6.4 (8.2 ± 0.1)	nd	73 (7.1 ± 0.4)
Partial agonists								
Pentazocine	10 (8.0 ± 0.1)	4.3 (8.4 ± 0.1)	8.6 (8.1 ± 0.1)	5.1 (8.3 ± 0.1)	3.3 (8.5 ± 0.1)	1.0 (9.0 ± 0.1)	nd	nd
Buprenorphine	2.1 (8.6 ± 0.2)	0.7 (9.1 ± 0.2)	1.0 (9.0 ± 0.3)	1.8 (8.7 ± 0.2)	2.0 (8.9 ± 0.3)	0.4 (9.3 ± 1.6)	nd	nd
Prototypic agonists								
Oxycodone	486 (6.3 ± 0.1)	278 (6.6 ± 0.2)	565 (6.3 ± 0.1)	286 (6.5 ± 0.2)	146 (6.8 ± 0.1)	39 (7.4 ± 0.2)	5300 (5.3 ±0.2)	2061 (5.7 ± 0.1)
Carfentanil	0.30 (9.5 ± 0.1)	0.21 (9.6 ± 0.1)	0.43 (9.4 ± 0.1)	0.21 (9.7 ± 0.1)	0.22 (9.8 ± 0.1)	0.14 (10.2 ± 0.1)	0.20 (9.7 ±0.2)	0.32 (9.5 ± 0.2)
Fentanyl	12.7 (7.9 ± 0.1)	3.4 (8.4 ± 0.2)	19.4 (7.7 ± 0.2)	6.8 (8.2 ± 0.1)	4.9 (8.3 ± 0.1)	0.9 (9.0 ± 0.1)	10.5 (8.0 ± 0.1)	7.8 (8.1 ± 0.q)
Morphine	11.9 (8.0 ± 0.1)	10.4 (8.0 ± 0.2)	7.2 (8.1 ± 0.2)	6.9 (8.2 ± 0.1)	5.5 (8.3 ± 0.1)	2.4 (8.6 ± 0.1)	205 (6.7 ± 0.1)	118 (6.9 ± 0.2)
Bitopics								
C3-guano	1234 (5.9 ± 0.2)	562 (6.2 ± 0.2)	1811 (5.7 ± 0.2)	505 (6.3 ± 0.2)	338 (6.5 ± 0.1)	305 (6.5 ± 0.2)	n.d.	23571(4.6 ± 0.5)
C5-guano	453 (7.3 ± 0.1)	79 (7.1 ± 0.1)	350 (6.5 ± 0.2)	23 (7.6 ± 0.1)	24 (7.6 ± 0.2)	8.8 (8.0 ± 0.1)	1210 (5.9 ± 0.2)	1760(5.7 ± 0.3)
C6-guano	132 (6.9 ± 0.2)	58 (7.2 ± 0.2)	33 (7.5 ± 0.3)	54 (57.3± 0.3)	72 (7.1 ± 0.2)	25 (7.6 ± 0.3)	n.d.	5191(5.3±0.6)
C7-guano	531 (6.3 ± 0.2)	450 (6.3 ± 0.2)	833 (6.1 ± 0.4)	81 (7.1 ± 0.2)	63 (7.2 ± 0.2)	161 (6.8 ± 0.1)	n.d.	n.d.
C9-guano	93.1 (7.0 ± 0.1)	60 (7.2 ± 0.2)	181 (6.7 ± 0.1)	27 (7.6 ± 0.1)	43 (7.6 ± 0.1)	27 (7.6 ± 0.1)	3215 (5.5 ± 0.2)	4089 (5.4 ± 0.2)
C11-guano	4745 (5.3 ± 0.1)	1311 (5.9 ± 0.3)	1065 (6.0 ± 0.3)	1242 (5.9 ± 0.1)	2038 (5.7 ± 0.1)	861 (6.1 ± 0.1)	n.d.	30885 (4.5 ± 0.2)

Potency [EC$_{50}$ nM (pEC$_{50}$ ± SEM)] are reported as estimates from simultaneous curve fitting of all biological replicates and include standard error.

Article

Extended Data Table 4 | Efficacy table for drugs profiled in Extended Data Fig. 6

Efficacy

Compounds	G_{i1} E_{max} % ± SEM	G_{i2} E_{max} % ± SEM	G_{i3} E_{max} % ± SEM	G_{oa} E_{max} % ± SEM	G_{ob} E_{max} % ± SEM	G_z E_{max} % ± SEM	β-arr1 E_{max} % ± SEM	β-arr2 E_{max} % ± SEM
Opioid Peptides								
DAMGO	99 ± 2	98 ± 3	98 ± 2	99 ± 2	100 ± 2	100 ± 2	100 ± 3.1	99± 2.8
Endomorphin-1	93 ± 1.6	95 ± 1.7	96 ± 1.4	109 ± 2.7	91 ± 2.6	105 ± 1.1	78 ± 3.1	82 ± 2.5
Leu-Enkephalin	92 ± 5.5	85 ± 7.2	93 ± 4.7	96 ± 3.8	102 ± 4.5	103 ± 5.2	100 ± 6.2	97 ± 6.8
Met-Enkephalin	92 ± 4.6	84 ± 6.6	103 ± 5.6	97 ± 3.7	96 ± 3.1	96 ± 3.1	102 ± 4.9	92 ± 4.5
Beta-Enkephalin	87 ± 3.6	102 ± 6.1	99 ± 2.8	91 ± 3.2	99 ± 5.9	104 ± 4.5	76 ± 4.8	90 ± 5.3
Dynorphin 1-17	92 ± 2.3	86 ± 3.8	90 ± 2.4	94 ± 4.4	92 ± 3.0	103 ± 1.4	50 ± 15	42±1.03
Biased Opiates								
PZM21	72 ± 2.0	84 ± 3.0	68 ± 1.9	92 ± 4.0	84 ± 3.1	98 ± 1.6	38 ± 3.3	24 ± 3.7
7-OH	31 ± 4.9	53 ± 5.1	nd	65 ± 4.9	56 ± 3.5	73 ± 2.8	nd	nd
TRV130	58 ± 3.8	58 ± 4.2	45 ± 3.3	67 ± 5.3	76 ± 4.4	98 ± 2.2	nd	26 ± 5.9
Partial agonists								
Pentazocine	75 ± 1.8	78 ± 3.8	73 ± 2.3	89 ± 4.0	89 ± 3.2	103 ± 1.5	7±4.51	13±4.04
Buprenorphine	59 ± 3.6	60 ± 3.7	38 ± 4.0	70 ± 4.3	68 ± 3.3	77 ± 3.5	nd	nd
Prototypic agonists								
Oxycodone	82 ± 3.2	83 ± 5.7	75 ± 3.8	102 ± 5.2	110 ± 3.4	102 ± 4.9	81 ± 7.9	80 ± 3.9
Carfentanil	97 ± 1.8	99 ± 2.2	102 ±1.8	119 ± 3.8	96 ± 4.1	112 ± 1.7	96 ± 7.0	97 ± 6.7
Fentanyl	105 ± 4.8	89 ± 6.2	92 ± 6.5	91 ± 4.3	100 ± 4.2	90 ± 3.4	92 ± 2.2	97 ± 3.4
Morphine	99 ± 4.1	86 ± 5.5	79 ± 6.2	93 ± 3.9	96 ± 3.2	98 ± 4.6	29 ± 2.0	37 ± 4.2
Bitopics								
C3-guano	71 ± 6.3	90 ± 6.5	70 ± 8.5	78 ± 5.1	83 ± 3.6	82 ± 5.3	n.d.	37 ± 11
C5-guano	80 ± 3.6	90 ± 3.7	85 ± 5.5	92 ± 2.3	87 ± 3.7	86 ± 3.2	32 ± 3.0	38 ± 5.3
C6-guano	71 ± 7.7	79 ± 7.2	56 ± 6.1	52 ± 7.0	64 ± 4.1	50 ± 5.3	n.d.	23±7.6
C7-guano	26 ± 1.8	38 ± 2.6	31 ± 5.2	45 ± 2.7	38 ± 2.6	73 ± 2.8	n.d.	n.d.
C9-guano	83 ± 3.4	96 ± 5.5	98 ± 4.6	95 ± 4.0	96 ± 2.8	99 ± 4.2	29 ± 2.7	61 ± 7.0
C11-guano	41 ± 2.5	45 ± 3.7	43 ± 5.2	54 ± 3.4	67 ± 3.2	82 ± 3.2	n.d.	80 ± 12

Efficacy (E_{max}% ± SEM) are reported as estimates from simultaneous curve fitting of all biological replicates and include standard error.

nature portfolio

Corresponding author(s): Georgios Skiniotis, Vsevolod Katritch, Brian Kobilka and Susruta Majumdar

Last updated by author(s): 02/21/2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	NMR spectra were recorded on Varian 400 MHz instrument as indicated in the supplementary information and collected via the Bruker Topspin Software (Bruker Topsin 3.5 pl 6). cryoEM data were collected using serialEM 3.8beta.
Data analysis	Graphpad Prism ver. 8 was used for all statistical analysis for the in vitro and in vivo assay experiments . Graphs were fit using a 3-[parameter logistics equation (nonlinear regression to fit log(dose) vs. response). For couple of panels Sigmaplots version 14.5 was used. NMR spectra were analyzed using Mestrenova 14.2.0 software. cryoEM data processing was done in RELION3.1. Model building was done using UCSF Chimera1.17, Phenix1.19.2, and Coot0.9.3. Ligand models were optimized by GemSpot-Maestro 2021-4. Figures were made using UCSF Chimera1.17 and PyMOL2.5.2. MD's used GROMACS software package version 2020.2. All the above molecular modeling operations were performed in ICM-Pro v3.8-5 molecular modeling package. G*Power 3.19.2 was used to calculate a minimum number of mice per testing condition.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data generated, analyzed in this study, methods used are included in the Supplementary Information. The cryo-EM maps and corresponding coordinates have been deposited in the Electron Microscopy Data Bank (EMDB) and the Protein Data Bank (PDB), respectively, under the following accession codes: EMD-26314 and 7U2L (C5-guano-uOR-Gi-scFv16), EMD-26313 and 7U2K (C6-guano-uOR-Gi).

The authors declare that all the data supporting the findings of this study are available within the article, Extended data and Supplementary Information files which contains synthetic procedures, NMR list of peaks for both novel and already described compounds, additional functional data at rodent and human receptors, biological protocols for receptor binding, activity and antinociception assays. All raw data for all figures is available in the source data files. The crystal structure of the active mu-opioid receptor was obtained from RCSB PDB using entry ID 5C1M. All compounds can be made available on reasonable requests from the authors.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](#)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Experiments were performed in biological triplicate (technical duplicates per biological replicate) which is standard for the field. Due to the amount of compounds tested and plate size limitations (ie, compounds per plate, number of cells per transfection, additional compounds being tested), our reference compound (DAMGO), had more biological replicates since it was run on every plate to allow for normalization across plates/replicates. Similarly in some cases the control morphine may have more replicates.
	Number of technical replicates and biological replicates are reported in the figure and table legends. Sample size was determined based on variability of the response deviating from the mean as indicated by the standard error of the mean (SEM), which is also represented in the figures. Aiming for 95% power with an alpha significance of p=0.05, and an ANOVA analysis with repeated measures within-between interaction, G*Power 3.19.2 was used to calculate a minimum number of mice per testing condition. For analgesic testing conditions for tail-withdrawal ("tail-flick") a large effect size (f=1.7) and eight measurements are assumed, resulting in an estimated sample size of 8 mice, with a predicted critical F-value of 1.95 and actual power = 1.0. For the inflammatory pain assay under these criteria, a small effect size (f=0.6) and 9 measurements are assumed of three doses, resulting in an estimated sample size of 10 mice, with a predicted critical F-value of 1.86 and actual power = 0.970. For assessment of visceral/chemical pain with the acetic acid writhing test (and using a one-way ANOVA to assess effects of large effect size (f=2.6) over four treatments), G-power calculates an estimated sample size of 10 mice, with a predicted critical F-value of 6.59 and actual power = 0.953. For anti-allodynic testing conditions for the CCI noxious stimuli under these criteria, a moderate-to-large effect size (f=1.2) and five measurements of three compound doses (and one control) are assumed, resulting in an estimated sample size of 10 mice, with a predicted critical F-value of 2.18 and actual power = 0.999. For respiratory measures, and with an assumed small effect size (f=0.45) over 18 measurements (corresponding to a minimum of 90 minutes), an estimated sample size of 12 mice will be needed for the respiratory/locomotor CLAMS testing, with a predicted critical F-value of 1.457 and actual power = 0.951. Assuming a small effect size (f=0.5) and two measures, G*Power estimated 24 mice per condition for conditioned place preference will be needed to achieve statistical significance in this work, defined with a critical F-value of 3.098, with actual power = 0.973. Much in the same manner, the operant model of neuropathic pain assumes a smaller effect size (f=0.42) and two measures, with and estimated 28 mice per condition needed to achieve statistical significance in this work, defined with a critical F-value of 3.40, with actual power = 0.964. Note that data was collected by cohort and analyzed after each experiment.
Data exclusions	No data were excluded for this study.
Replication	Data were replicated using technical and independent replicates. See figure and table legends for specific details.
Randomization	For the behavioral studies, mice were randomly assigned across treatments. For functional and binding assays, randomization is not relevant as no group allocations were performed.
Blinding	The in vitro screening of analogs was done unblinded and then to confirm efficacy and receptor selectivity we ran blinded experiments in two independent labs. The analgesia experiments done on male C57BL/6 mice (22–30 g) and CD1 mice (28-35 g) were performed by blinding the experimenter to drug identity. Note that the analgesia experiments done on the wild-type and MOR KO mice were also done by blinding the experimenter to strain type as well as drug.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☐ ☒	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☐ ☒	Animals and other organisms
☒ ☐	Human research participants
☒ ☐	Clinical data
☒ ☐	Dual use research of concern

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T (CRL-3216) was purchased on 06/08/2017 from American Type Culture Collection (Rockville, MD, USA). CHO-K1 cells were obtained from the American Type Culture Collection American Type Culture Collection (Rockville, MD, USA). Sf9 and Hi5 insect cells were purchased from Expression Systems.
Authentication	Cells have not been authenticated after purchase
Mycoplasma contamination	Cells have not been tested for mycoplasma
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in the study

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Male C57BL/6J mice (24–38 g, 8-12 weeks) were purchased from Jackson Laboratories (Bar Harbor, ME). Male CD1 mice (29-45 g, 8–12 weeks of age) were purchased from Charles River Laboratories (Wilmington, MA). Male MOR KO were bred in the McLaughlin laboratory at University of Florida and used when 8-12 weeks old. Progenitors of the colonies for µOR KO were obtained from Jackson Labs. All mice used throughout the manuscript were opioid naïve. All mice were maintained on a 12 hr light/dark cycle with Purina rodent chow and water available ad libitum and housed in groups of five until testing. These mice were kept at a constant temperature of 22±2°C, and relative humidity was maintained at 40–50%.
Wild animals	Study did not involve wild animals.
Field-collected samples	Study did not involve collected samples collected from field.
Ethics oversight	This study was carried out in accordance with the recommendations of the 2011 NIH Guide for the Care and Use of Laboratory Animals and ARRIVE guidelines overseen by the Institutional Animal Care and Use Committee at the University of Florida. The protocol was preapproved by the Institutional Animal Care and Use Committee at the University of Florida. Treatment groups were blinded and randomly assigned animals. Protocol number (#202009530) approved by IACUC was used.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

cancers

MDPI

Article

Defining A Liquid Biopsy Profile of Circulating Tumor Cells and Oncosomes in Metastatic Colorectal Cancer for Clinical Utility

Sachin Narayan [1], George Courcoubetis [1], Jeremy Mason [1,2,3], Amin Naghdloo [1], Drahomír Kolenčík [4], Scott D. Patterson [5], Peter Kuhn [1,2,3,6,7,8,*] and Stephanie N. Shishido [1]

[1] Michelson Center for Convergent Bioscience, Convergent Science Institute in Cancer, Dornsife College of Letters, Arts and Sciences, University of Southern California, Los Angeles, CA 90089, USA
[2] Catherine & Joseph Aresty Department of Urology, Institute of Urology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA
[3] Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA
[4] Biomedical Center, Faculty of Medicine in Pilsen, Charles University, 32300 Pilsen, Czech Republic
[5] Gilead Sciences, Inc., Lakeside Drive, Foster City, CA 94404, USA
[6] Department of Biomedical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089, USA
[7] Department of Biological Sciences, Dornsife College of Letters, Arts, and Sciences, University of Southern California, Los Angeles, CA 90089, USA
[8] Department of Aerospace and Mechanical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089, USA
*   Correspondence: pkuhn@usc.edu; Tel.: +1-213-821-3980

Simple Summary: Metastatic colorectal cancer (mCRC) is typified by its tumor heterogeneity and changing disease states, suggesting that personalized medicine approaches could be vital to improving clinical practice. As a minimally invasive approach, the liquid biopsy has the potential to be a powerful longitudinal prognostic tool. We investigated mCRC patients' peripheral blood samples using an enrichment-free single-cell approach to capture the broader rare-event population beyond the conventionally detected epithelial-derived circulating tumor cell (CTC). Our analysis reveals a heterogenous profile of CTCs and oncosomes not commonly found in normal donor samples. We identified select rare cell types based on their distinct immunofluorescence expression and morphology across multiple assays. Lastly, we highlight correlations between enumerations of the blood-based analytes and progression-free survival. This study clinically validates an unbiased rare-event approach in the liquid biopsy, motivating future studies to further investigate these analytes for their prognostic potential.

Abstract: Metastatic colorectal cancer (mCRC) is characterized by its extensive disease heterogeneity, suggesting that individualized analysis could be vital to improving patient outcomes. As a minimally invasive approach, the liquid biopsy has the potential to longitudinally monitor heterogeneous analytes. Current platforms primarily utilize enrichment-based approaches for epithelial-derived circulating tumor cells (CTC), but this subtype is infrequent in the peripheral blood (PB) of mCRC patients, leading to the liquid biopsy's relative disuse in this cancer type. In this study, we evaluated 18 PB samples from 10 mCRC patients using the unbiased high-definition single-cell assay (HDSCA). We first employed a rare-event (Landscape) immunofluorescence (IF) protocol, which captured a heterogenous CTC and oncosome population, the likes of which was not observed across 50 normal donor (ND) samples. Subsequent analysis was conducted using a colorectal-targeted IF protocol to assess the frequency of CDX2-expressing CTCs and oncosomes. A multi-assay clustering analysis isolated morphologically distinct subtypes across the two IF stains, demonstrating the value of applying an unbiased single-cell approach to multiple assays in tandem. Rare-event enumerations at a single timepoint and the variation of these events over time correlated with progression-free survival. This study supports the clinical utility of an unbiased approach to interrogating the liquid

---

biopsy in mCRC, representing the heterogeneity within the CTC classification and warranting the further molecular characterization of the rare-event analytes with clinical promise.

Keywords: liquid biopsy; rare cell; circulating tumor cells; oncosomes; colorectal cancer; heterogeneity; multi-assay; high-definition single-cell assay

1. Introduction

Colorectal cancer (CRC) is the world's third most common cancer and second leading cause of oncology-related deaths [1]. Most notably, CRC solid tumors are marked by their extensive cellular heterogeneity and proliferation owed to the rapid rate of epithelial self-renewal in the intestines [2–4]. The variety of tumor microenvironments, genetic mutations, and disease subtypes suggests that real-time individualized analysis and subsequent clinical decision-making could improve patient outcomes [5–7].

As a minimally invasive procedure, the liquid biopsy has the potential to be that critical element in the longitudinal evaluation of CRC by characterizing the disease's pathophysiology and mechanisms of metastasis [8–10]. Much of the current liquid biopsy analysis focuses on two primary biomarkers: circulating tumor cells (CTCs) and cell-free DNA (cfDNA). Additionally found in the blood, carcinoembryonic antigen (CEA) is one of the foremost prognostic hallmarks of CRC [11,12]. Previous investigations have correlated CTC counts to CEA levels, with the two used in conjunction to accurately predict survival outcomes [13]. Generally, a higher number of CTCs indicates poorer patient outcomes in CRC [14,15], although nuances emerge when considering morphologically defined CTC subtypes and the change in cell populations over time [16,17]. With the variety of detectable biomarkers in circulation, the liquid biopsy could aid in tackling some of the clinical challenges of CRC.

Previously, CellSearch® (Menarini, Raritan, NJ, USA) was the first platform to receive regulatory approval in CRC via a 510(k) clearance for the enumeration of CTCs to monitor metastatic colorectal, breast and prostate cancer [18]. The platform uses an enrichment-based methodology for the detection of a singular type of CTC defined by the expression of epithelial cell adhesion molecule (EPCAM) and cytokeratin (CK) without the expression of CD45 [19], thereby limiting the liquid biopsy field's current understanding of CTC. CellSearch® uses a uniform threshold for CTC positivity at any timepoint during the patient's treatment. In CRC, this threshold is 3 CTCs/7.5 mL of blood [20]. CellSearch® has shown that a higher frequency of CTCs is associated with poorer overall survival (OS) and progression-free survival (PFS) in CRC [21,22]. However, CellSearch® and similar systems are not commonly utilized by clinicians treating CRC [23,24]. This limited utility could be attributed to a lack of timepoint-specific standards [18] and infrequent CTC kinetics analysis, despite its clinical promise [16]. Most early-generation liquid-biopsy platforms employ enrichment-based approaches that detect a limited CTC population in the peripheral blood (PB) of CRC patients [25–27] and overlook other cellular subtypes with nuanced survival implications [16,28]. In a cancer type known for its tumor heterogeneity like CRC, enrichment-based approaches limit the liquid biopsy's potential clinical utility, thereby warranting an unbiased single-cell approach that focuses on all rarity in the bloodstream.

This study utilized the third-generation high-definition single-cell assay (HDSCA3.0), which is a validated "no cell left behind" immunofluorescence (IF) assay that detects and characterizes all rare events from the liquid biopsy [25,28–33]. Commercialized by Epic Sciences, it has demonstrated clinical utility as a predictive marker in prostate cancer [34–36]. Furthermore, it allows for downstream genomic and proteomic analysis [37,38] and adheres to the standards of the Blood Profiling Atlas Commons [39]. A prior investigation into a cohort of metastatic CRC (mCRC) patients with HDSCA2.0 revealed a 35% CTC positivity rate [25], comparable to CellSearch® positivity rates among similar cohorts [30,37,40]. A subsequent study with the same platform highlighted the importance of CTC sub-

types, time of sample collection and changes in cellular populations during treatment in understanding the value of the liquid biopsy in mCRC patient care [16]. Beyond CTC subtypes, prior studies have identified other rare cells such as circulating endothelial cells (CECs) [41,42] with prognostic implications in mCRC. In addition to rare cellular analytes, tumor-derived oncosomes and extracellular vesicles have been shown to promote tumori-genesis and chromosomal deletion across cancer types [43–45]. Now, the third generation of HDSCA detects a heterogeneous CTC and oncosome population with various surface biomarkers and an unbiased computational methodology for the detection of epithelial, mesenchymal, endothelial and immune cells [46,47]. Herein, we analyzed 18 PB samples from 10 mCRC patients using two IF protocols to represent a comprehensive CTC and onco-some liquid-biopsy profile, highlighting previously unidentified rare events and correlating analytes to patient outcomes.

2. Materials and Methods

2.1. Study Design

This study includes a total of 18 PB samples collected between May 2016 to March 2017 from 10 patients with mCRC. Patients were found as part of the GS-US-296-0101 phase I clinical trial (#NCT01803282) evaluating the safety and tolerability of a novel therapeutic in combination with standard-of-care chemotherapy in two different mCRC indications. Apart from their diagnosis of CRC and survival data, no other clinical or demographic information was available for this study cohort per IRB protocol at the time of enrollment. Patients 1, 3, 5, 6 and 9 were first-line inoperable mCRC patients receiving the test compound in combination with mFOLFOX6 and bevacizumab. Patients 2, 4, 7, 8 and 10 were second-line inoperable patients receiving the test compound in combination with FOLFIRI and bevacizumab. PB samples were collected on cycle 1, day 1 and cycle 3, day 1 of therapy, referred to as Draw 1 and Draw 2, respectively. All patients progressed, and PFS was provided for 9 of the 10 patients. In addition, PB samples from 50 normal donors (ND) with no known pathology were collected and provided by Epic Sciences (San Diego, CA, USA).

2.2. Blood Processing

PB samples were collected in 10 mL collection tubes (Cell-free DNA, Streck, La Vista, NE, USA) and were processed as previously described [30,47]. As a brief synopsis, after red blood cell lysis, the nucleated cell fraction was plated as a monolayer of ~3 million cells per slide (Marienfeld, Lauda, Germany) before cryobanking at −80 °C.

2.3. IF Staining Protocols

Samples were analyzed with the previously described workflow for high-resolution imaging and the characterization of tumor cells at a single-cell level [30]. Slides were stained by the IntelliPATH FLX[TM] autostainer (Biocare Medical LLC, Irvine, CA, USA) in batches of 50. Our validated HDSCA protocols utilize a cocktail of pan-cytokeratin (CK), CD45 antibodies and DAPI [30,31]. In further detail, samples were fixed with a 2% neutral buffered formalin solution (VWR) for 20 min followed by permeabiliza-tion using 100% cold methanol for 5 min and blocking nonspecific binding sites with 10% goat serum (Millipore) for 20 min. This is followed by an antibody cocktail consisting of mouse IgG1/Ig2a anti-human cytokeratins 1, 4, 5, 6, 8, 10, 13, 18 and 19 (clones: C-11, PCK-26, CY-90, KS-1A3, M20, A53-B/A2, C2562, Sigma, St. Louis, MO, USA), mouse IgG1 anti-human cytokeratin 19 (clone: RCK108, GA61561-2, Dako, Carpinteria, CA, USA) and mouse anti-human CD45:Alexa Fluor® 647 (clone: F10-89-4, MCA87A647, AbD Serotec, Raleigh, NC, USA). To complete the staining, slides were incubated with Alexa Fluor® 555 goat anti-mouse IgG1 antibody (A21127, Invitrogen, Carlsbad, CA, USA) and 4′,6-diamidino-2-phenylindole (DAPI; D1306, ThermoFisher) prior to being mounted with a glycerol-based aqueous mounting media [48].

Two distinct IF protocols were applied: 1) Landscape and 2) CDX2-targeted. The Land-scape protocol identifies epithelial, mesenchymal, endothelial and immune cells through

the addition of 100 ug/mL of a goat anti-mouse IgG monoclonal Fab fragments (115-007-003, Jackson ImmunoResearch, West Grove, PA, USA), rabbit IgG anti-human vimentin (Vim) (clone: D21H3, 9854BC, Cell Signaling, Danvers, MA, USA) as a fourth color and mouse IgG1 anti-human CD31:Alexa Fluor® 647 mAb (clone: WM59, MCA1738A647, BioRad, Hercules, CA, USA) to the CD45 channel, hereafter referred to as CD45/CD31 [47]. The CDX2-targeted protocol utilizes the colon-specific CDX2 monoclonal antibody EPR2764Y (Abcam, Cambridge, UK) as a fourth color for further characterization [25]. CDX2 is a transcription factor expressed throughout the intestinal epithelium and has been effectively used as a marker for intestinal carcinomas [49]. The ND samples were stained with the Landscape protocol.

2.4. Scanning and Analysis

Slides were imaged at 2304 frames per slide using automated high-throughput fluorescence scanning microscopy at 100× magnification with exposures and gain set to yield the same background intensity level for normalization purposes. The numeration of cell classifications was converted to concentration based on the sample leukocyte concentration measured at processing and the number of DAPI-positive nuclei detected. White blood cell (WBC) counts in the PB sample were determined automatically (Medonic M-series Hematology Analyzer, Clinical Diagnostic Solutions Inc., Fort Lauderdale, FL, USA). The number of WBCs per slide was utilized in the calculation of the exact amount of blood analyzed, leading to rare-event enumerations presented in events/mL. Cells of interest were further imaged at higher magnification (400×). IF signal expression is categorized as filamentous, diffuse or punctate, as previously described [50].

2.5. Rare Event Detection Approach

Cells were identified via a rare-event detection method termed OCULAR [47,48]. This algorithm uses feature extraction, principal component analysis (PCA) and hierarchical clustering on the principal components to achieve four distinct tasks: fluorescent image feature extraction (761 parameters), rare-event detection (distinguishing between common and rare DAPI-positive events and DAPI-negative events), rare-cell classification and report generation. Image analysis was performed as previously reported [47]. In brief, all events were segmented to generate nuclear and/or cytoplasm masks for feature extraction, which was followed by a dimensionality reduction using principal components and hierarchical clustering to separate common cells (mainly WBCs) and rare cells in each image frame. The manual classification of rare events into CTC or oncosome subgroups was conducted based on biomarker expression in the four fluorescence channels for each IF protocol. Classifications were validated by multiple hematopathologist-trained technical analysts. Previously described as large extracellular vesicles in the context of HDSCA [46,48,51], oncosomes were identified as circular, DAPI-negative events with positive CK expression [44,52,53]. Initially presented as adjacent to nucleated common cells or as individual DAPI-negative events, oncosomes were manually classified and confirmed by trained analysts. The nomenclature for channel-type classifications utilizes those positive channels (Landscape example: CK|Vim = DAPI-positive, CK-positive, Vim-positive, CD45/CD31-negative). Oncosome channel-type classifications are preceded by the abbreviation "Onc" (Landscape example: Onc CK|Vim = DAPI-negative, CK-positive, Vim-positive, CD45/CD31-negative). While rare cell types are predominantly referred to by their channel-type classification, two specific CTC populations, epithelial CTCs (Epi.CTC) and mesenchymal CTCs (Mes.CTC), are evaluated in this study. As previously described [48], Epi.CTCs are CK-positive, Vim-negative and CD45/CD31-negative with a clearly defined nucleus (DAPI). Mes.CTCs are CK-positive, Vim-positive and CD45/CD31-negative and have clearly defined nuclei.

2.6. Multi-Assay Analysis

With sample-matched slides stained by both IF protocols (Landscape and CDX2-targeted), OCULAR's uniform examination of the rare events allowed for a multi-assay

analysis. OCULAR derives 761 morphological parameters from IF $100\times$ magnification images of detected cellular events. To perform the multi-assay analysis, we selected a subset of 8 morphometric parameters, consisting of the most representative features. With both IF protocols containing DAPI and CK, the median intensity for these two channels were chosen. In addition, we included the eccentricity and area of both the cell and nucleus, the ratio of nuclear to cellular area and the average distance of the cell outline to the center of the nucleus. Using these 8 shared morphometric features, 5661 rare cells across both assays were grouped together using a hierarchical clustering model. An agglomerative clustering algorithm was used, imported from the scikit-learn library version 0.23.2 [54] in Python. We used a Euclidian metric to compute the distance and the ward linkage criterion. In addition, 111 cells and the oncosome population were removed after manual inspection from the clustering due to highly aberrant nuclear and membrane masking. The cohort's rare-cell population with various cluster assortment options are included as part of an interactive webpage that can be used for additional analyses and discovery (https://pivot.usc.edu/pivot/CRC_MultiAssay.html). The optimal number of clusters from 2 to 16 was selected based on quantitative cluster-separation metrics and the separation of noteworthy, rare cell types. Based on the silhouette average method, 8 clusters provided the optimal separation between the cell groups. When increasing the number of clusters from 8 through 12, cluster 1 is the primary group undergoing rearrangement, but subsequent cluster combinations do not perform as well in the silhouette average metric or in their division of rare cell types of interest. Alongside the clustering approach, a Spearman's rank correlation [55] analysis was performed between the rare-event enumerations of the two IF protocols.

2.7. Survival Analysis

To perform survival analysis on the 9 patients with known progression time, the Spearman's rank correlation coefficient [55] was calculated for all liquid-biopsy analytes versus the PFS. Then, only liquid-biopsy analytes with statistically significant entries were used for subsequent visualization. To visualize the statistically significant entries, the Kaplan–Meier (KM) curves [56] for the patients' PFS were plotted. Then, to depict the effect of a statistically significant liquid-biopsy analyte on PFS, the patients were stratified based on their respective counts per milliliter of blood for the given liquid-biopsy analyte. The stratification was done by using the median counts per milliliter of blood as a threshold, separating the patients into two groups. Finally, the two patient subgroups were plotted together with the original KM curve of the population. The statistical analysis was done using scipy [57], and the KM curves were plotted using scikit-survival library [58] in Python.

2.8. Statistical Analysis

The distinction between the rare-event enumerations of the mCRC and ND samples was determined by a Wilcoxon rank sum test [59]. Statistical correlations between rare-event enumerations were performed using the Spearman's rank correlation coefficient [55]. A correlation was significant if the two-tailed p-value ≤ 0.05. The statistical analysis was done using scipy [57] in Python.

3. Results

3.1. Landscape Rare-Event Detection: Rare Cells and Oncosomes

PB samples were stained with the Landscape IF protocol and analyzed by OCULAR to identify the rare events with biomarkers highlighting epithelial, endothelial, mesenchymal and immune cell origin. Rare-event frequencies, enumerations and sample positivity (≥ 5 events/mL) from the Landscape and CDX2 IF protocols are reported in Table 1. For the 18 mCRC samples analyzed, on average 0.53 (standard error 0.05, median 0.53, range 0.22–0.97) mL of PB was used for 1 test, thus the sensitivity of the analysis is limited by the blood volume characterized. A representative subset of the rare-cell and oncosome

populations detected in the Landscape-stained samples is displayed in Figure 1A, with enumerations, frequencies and comparisons to the ND cohort, and select morphometrics provided in Figure 1B, 1C and 1D, respectively. The sample enumeration of the ND cohort is depicted in Supplemental Figure S1.

Table 1. Rare-event frequencies, enumerations and sample positivity from the Landscape and CDX2-targeted immunofluorescence (IF) protocols. The sample positivity threshold of ≥ 5 events/mL was determined by comparisons to the rare-event enumerations of a randomly selected normal donor cohort. The frequency of each classification is provided as a percentage of the total rare-event profile for each IF protocol.

IF Protocol	Event Classification	Sample Positivity	Mean (Events/mL)	Standard Error (±Events/mL)	% of Total Rare Events	Median (Events/mL)	Range (Events/mL)
Landscape	DAPI only	15/18	13.69	2.9	3.55	10.84	0.00–51.57
	CK (Epi.CTC)	18-Jul	57	36.36	14.77	3.55	0.00–549.63
	Vim	18-Nov	6.64	1.27	1.72	7.3	0.00–20.16
	CD45/CD31	18-Oct	10.29	2.44	2.67	8.05	0.00–33.16
	CK|Vim (Mes.CTC)	18-Jun	7.6	1.4	1.97	1.4	0.00–91.29
	CK|CD45/CD31	18-Apr	4.66	1.86	1.21	2.13	0.00–30.74
	Vim|CD45/CD31	18-Nov	24.52	8.73	6.35	7.79	0.00–121.00
	CK|Vim|CD45/CD31	15/18	100.4	35.02	26.01	12.89	0.00–453.13
	Onc CK	18-Dec	71.59	38.39	18.55	6.64	1.06–657.60
	Onc CK|Vim	18-Oct	32.72	13.64	8.48	7.38	0.00–217.65
	Onc CK|CD45/CD31	0/18	1.47	0.36	0.38	1.12	0.00–4.47
	Onc CK|Vim|CD45/CD31	14/18	55.37	16.22	14.35	34.32	0.00–268.32
CDX2-targeted	DAPI only	18/18	63.72	12.8	8.33	45.29	16.94–226.94
	CK	18-Dec	88.14	43.95	11.53	11.97	0.00–597.32
	CDX2	18-Dec	11.45	3.28	1.5	7.27	1.04–60.22
	CD45	18-Mar	5.44	3.35	0.71	0	0.00–59.22
	CK|CDX2 (CDX2.CTC)	14/18	19.95	6.71	2.61	11.34	0.00–124.08
	CK|CD45	18-Dec	36.66	13.71	4.79	10.02	0.00–203.63
	CDX2|CD45	18-Dec	29.37	12.46	3.84	9.8	0.00–185.68
	CK|CDX2|CD45	14/18	143.67	101.35	18.79	21.79	1.44–1843.34
	Onc CK	15/18	123.55	60.73	16.16	26	2.31–1035.65
	Onc CK|CDX2	18/18	222.4	65.04	29.08	114.87	15.55–1151.64
	Onc CK|CD45	0/18	0.06	0.06	0.01	0	0.00–1.06
	Onc CK|CDX2|CD45	18-Oct	20.27	7.64	2.65	14.5	0.00–1138.28

For all mCRC samples, total rare-event (total cells and oncosomes) detection had a median of 287.46 (mean 387.74 ± 75.33) events/mL. For ND samples, total rare-event detection had a median of 40.05 (mean 49.96 ± 4.18) events/mL. A significant difference was observed between the mCRC patients and ND ($p < 0.0001$; Figure 1D).

Rare cells comprised 58.25% of the total rare-event profile from the Landscape-stained mCRC samples. The rare cells detected in mCRC patient samples were highly heterogenous in their signal expression and morphology (Figure 1E,F). Total rare-cell detection for the mCRC samples had a median of 124.66 (mean 224.80 ± 51.55) cells/mL. The ND samples presented with a median rare-cell detection of 34.46 (mean 43.21 ± 3.94) cells/mL. A significant difference in total rare-cell detection was observed between the mCRC patients and ND samples ($p = 0.0112$).
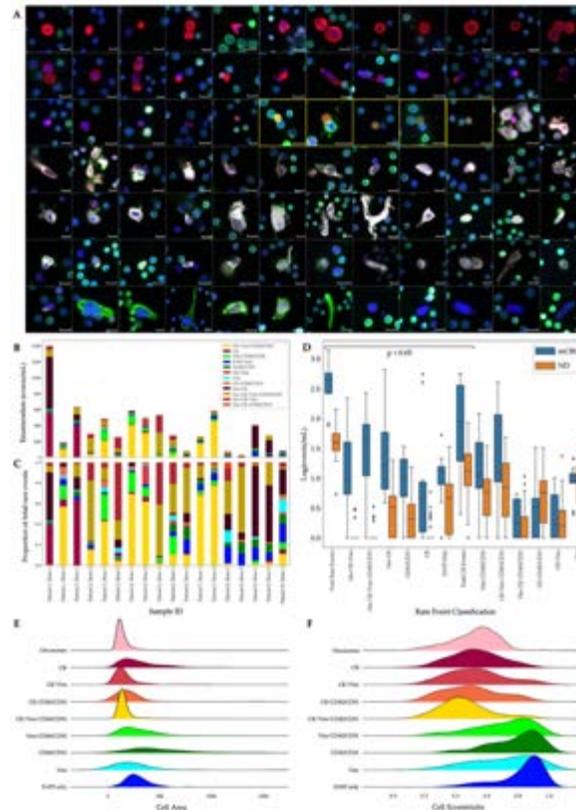
Figure 1. Landscape-stained samples analyzed by OCULAR. (**A**) Representative gallery from the metastatic colorectal cancer (mCRC) cohort showing the morphological heterogeneity of the detected rare events. DAPI: blue, cytokeratin (CK): red, Vim: white, CD45/CD31: green. Events from each of the 10 patients are represented in the gallery. The five oncosomes displayed are bordered by yellow boxes. Events are ordered by decreasing CK signal intensity. Images taken at 400× magnification. Scale bars represent 10 μm. (**B**) Rare-event enumeration (events/mL) and (**C**) frequency (%) per patient and draw. (**D**) Enumeration comparison of Draw 1 mCRC and normal donor samples ordered by statistical significance. Symbols indicate outliers. The first 8 classifications from the left are different between mCRC and normal donors ($p < 0.05$). (**E**) Cellular area and (**F**) cellular eccentricity per rare-event classification detected in the mCRC cohort.

Total CK-positive events were detected with a median of 45.41 (mean 169.66 ± 46.65) events/mL from all mCRC samples. The ND samples had a median of 12.39 (mean 18.96 ± 2.70) CK-positive events/mL. There was a statistically significant difference in total CK-positive event detection between the mCRC patients and ND samples ($p = 0.0070$). The total CK-positive cell population constituted 75.47% of all the rare cells. Epi.CTCs were detected with a median of 3.55 (mean 57.00 ± 36.36) cells/mL from mCRC patient samples, which is a significantly higher incidence compared to ND samples ($p = 0.0023$). Epi.CTCs were only identified in 7 of 18 mCRC samples, with 2 samples containing 92.58% of the total cohort's population. This speaks to the limited frequency of CK-only CTCs in the liquid biopsy of mCRC, here using approximately 0.5 mL of PB. Similarly, Mes.CTCs were only found in 6 of 18 samples and with a median of 1.40 (mean 7.60 ± 1.40) cells/mL in the mCRC cohort. There was not a significant difference in the Mes.CTC count between the mCRC and ND samples. The CK|Vim|CD45/CD31 cells were the most frequent rare cell across the mCRC cohort, but no statistically significance difference was detected between the mCRC and ND samples.

Other detectable rare cells in the mCRC samples included morphologically distinct CD45/CD31-only (median 8.05; mean 10.29 $\pm$ 2.44), DAPI-only (median 10.84; mean 13.69 $\pm$ 2.90), Vim-only (median 7.30; mean 6.64 $\pm$ 1.27) and Vim | CD45/CD31 (median 7.79, mean 24.52 $\pm$ 8.73) cells/mL. The DAPI-only cells were detected at a higher prevalence in mCRC samples compared to the ND samples (p = 0.0047).

Two morphologically distinct cell types were identified as unique subsets of their broader channel-type classifications. The first was the large, morphologically distinct CD45/CD31 cell population shown in Figure 2A. As a subset of the CD45/CD31 channel-type classification, this cell type was found in 9 of the cohort's 18 samples (Figure 2B). Their distinction highlights the importance of morphological analysis that goes beyond channel-type classifications. The cells of interest possess a punctate CD45/CD31 signal, which is distinct from the diffuse CD45/CD31 signal that typifies the surrounding WBCs. Notable morphometrics quantitatively distinguish this cell group from common WBCs with a larger cell size, larger nuclear size, and higher cellular eccentricity. Even within this category, there is heterogeneity, evidenced by the varying shapes and the nuclear-to-membrane ratios of the cells. These cells are hypothesized to be megakaryocytes.
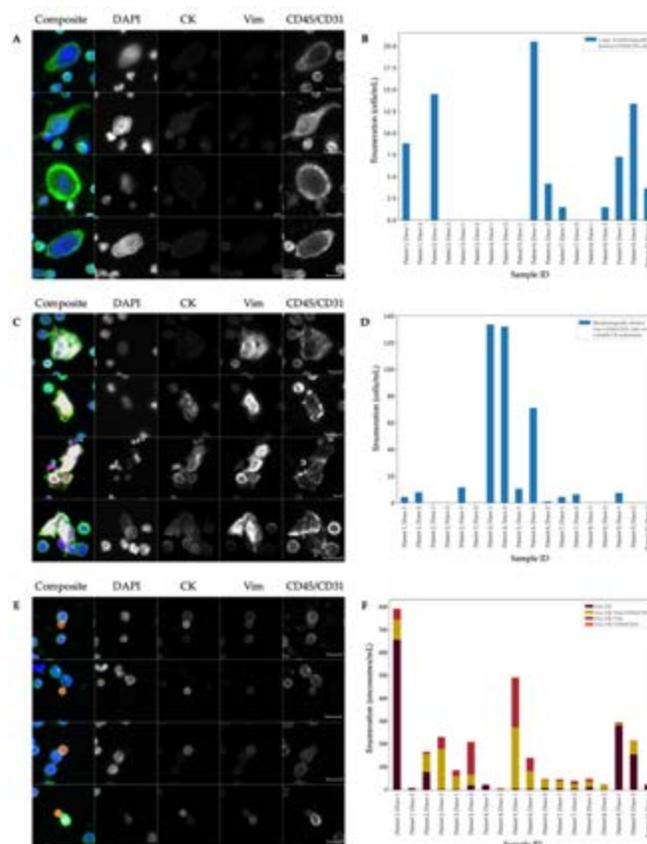

Figure 2. Select rare-event populations in the Landscape-stained metastatic colorectal cancer (mCRC) samples. (**A**) Panel gallery images of the large, morphologically distinct CD45/CD31 cells. (**B**) Large, morphologically distinct CD45/CD31 cell enumeration per patient and draw. (**C**) Panel gallery images of the morphologically distinct Vim | CD45/CD31 cells with variable cytokeratin (CK) expression. (**D**) Morphologically distinct Vim | CD45/CD31 cells with variable CK expression enumeration per patient and draw. (**E**) Panel gallery images of the oncosome population with differential signal expression (**F**) Oncosome enumeration by channel classifications per patient and draw. Images are taken at 400× magnification. Scale bars represent 10 μm.

The second rare-cell group of interest consists of morphologically distinct Vim | CD45/CD31 cells with variable CK expression, as shown in Figure 2C. These cells are found in 11 of 18 samples and across both draws (Figure 2D). The samples from Patient 4 contain the vast majority (81.37%) of the morphologically distinct Vim | CD45/CD31 cells detected in this cohort. The image analysis of these cells revealed a filamentous Vim signal along with a punctate CD45/CD31 signal. The variable CK expression in signal intensity and appearance suggests that there are additional subtypes within this cell type. A total of 45.60% of the total morphologically distinct Vim | CD45/CD31 cell population is CK-positive. Further observation highlights heterogeneous CK expression within the CK-positive subtype, epitomized by the punctate and filamentous signal on the second and third cell of Figure 2C respectively. Beyond the analysis of the IF signal, 35.95% of the morphologically distinct Vim | CD45/CD31 cells found in this cohort are found clustered near one another (bottom of Figure 2C). Additional key morphometrics that distinguish this cell population from surrounding WBCs include a large cell size and eccentric cellular membrane. These cells are hypothesized to be endothelial cells.

In addition to the rare-cell groups, OCULAR identified a sizeable population of oncosomes. Oncosomes accounted for 41.75% of the rare events in the Landscape-stained samples. Morphologically, these vesicles ranged up to the size of neighboring WBCs (~10 μm) and were present in the cellular fraction of blood after centrifugation. As Figure 2E depicts, these vesicles were found in contact with adjacent nucleated cells and in isolation, with 51.09% of this cohort's oncosome population belonging to the latter. Furthermore, their IF signal was diffuse, suggesting an evenly distributed expression across the vesicle. The oncosomes expressing CK were generally the most prevalent (Figure 2F). As the most common subtype, all 18 samples were positive for Onc CK with a median of 6.64 (mean 71.59 ± 38.39) events/mL. The Onc CK | Vim | CD45/CD31 was positive in 16 of 18 samples and had a median of 34.32 (mean 55.37 ± 16.22) events/mL. The Onc CK | Vim was also present in 16 of 18 samples, with a median of 7.38 (mean 32.72 ± 13.64) events/mL. The Onc CK | Vim | CD45/CD31 and Onc CK | Vim counts were found to be highly positively correlated ($p = 0.002$, $\tau = 0.68$).

In comparison to the mCRC and ND samples, 6 specific channel-type rare-event classifications were statistically distinct across the cohorts (Figure 1D). Three of the significantly different channel-type classifications were rare cells that were detected at a higher prevalence in mCRC samples compared to the ND: Epi.CTC ($p = 0.0023$), DAPI only ($p = 0.0494$) and CD45/CD31 ($p = 0.0004$). Three oncosome channel-type classifications were observed at greater numbers in the mCRC patient samples compared to the ND samples: Onc CK ($p = 0.0001$), Onc CK | Vim ($p < 0.0001$) and Onc CK | Vim | CD45/CD31 ($p < 0.0001$).

3.2. Analysis of the CDX2-Targeted Protocol

To complement the vast heterogeneity of rare events across the epithelial, mesenchymal and endothelial cell types presented in the Landscape protocol, the CDX2-targeted protocol was utilized to specifically interrogate and identify circulating rare events of colorectal origin. Slides from the same PB tubes were stained with the CDX2-targeted protocol and analyzed by OCULAR, allowing for a sample-matched study design across two IF assays. All rare-event frequencies, enumerations and sample positivity (≥ 5 events/mL) from the CDX2-targeted protocol are reported in Table 1. A representative subset of the rare-cell and oncosome populations are displayed in Figure 3A, with enumerations and frequencies of the channel-type classifications in Figure 3B,C. Further analysis of the CDX2-targeted cohort by HDSCA's first generation CK-focused approach is depicted in Supplemental Table S1, Figures S2 and S3.

Rare cells comprised 52.10% of the total rare-event profile from the CDX2-targeted protocol. From the OCULAR analysis, the total CK-positive cell population across all 18 samples had a median of 108.74 (mean 288.43 ± 119.17) cells/mL. The image analysis of the CK-positive cell groups revealed extensive heterogeneity between and within the channel-type classifications. The consistency of the IF signal across the cells of interest

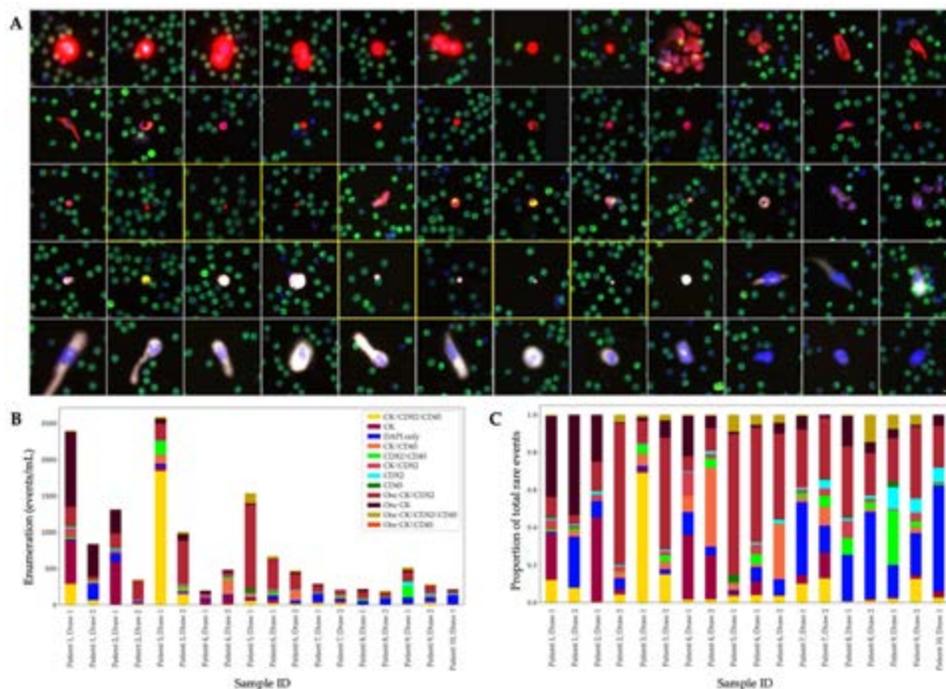varies within categories. Morphological differences were also observed between cells of the same channel type.

Figure 3. CDX2-targeted samples analyzed by OCULAR. (**A**) Representative rare-event gallery. DAPI: blue, cytokeratin (CK): red, CDX2: white, CD45: green. Events from each of the 10 patients are represented in the gallery. The eight oncosomes displayed are bordered by yellow boxes. Events are ordered by decreasing CK intensity. Images taken at $100\times$ magnification. Scale bar for all images is shown in the bottom right cell, representing 10 μm. (**B**) Enumeration (events/mL) of each channel classification per patient and draw. (**C**) Frequency (%) of each channel classification per patient and draw.

With the added CDX2 marker, we were able to evaluate CTCs for their potential colorectal origin. CDX2-positive CTCs or CDX2.CTCs (also referred to as CK|CDX2 in Table 1) were found in 14 of 18 samples, but not in high frequencies, only constituting 2.52% of the total rare-event profile. The first three images from the top left of Figure 3A are representative of this cell type, and, as the gallery shows, these CDX2.CTCs were also found clustered together.

Beyond the CK-positive populations, additional rare cell types were prominent across the CDX2-stained samples: DAPI-only, CDX2-only, CD45-only and CDX2|CD45 classifications (Table 1). The CK-negative rare-cell population detected by OCULAR was positive in all 18 samples, with a median of 87.13 (mean 109.98 ± 18.58) cells/mL. Among this group, a subset of large and eccentric cells with punctate CDX2 expression and sizeable nuclei (see the bottom row of Figure 3A) were found in 11 of the 18 samples. Lastly, all 18 of samples were positive for DAPI-only rare cells, with a median of 45.29 (mean 63.72 ± 12.80) cells/mL using the CDX2-targeted protocol.

3.3. Multi-Assay Analysis

A multi-assay comparison was conducted on samples stained both with the CDX2-targeted and Landscape IF protocols. An analysis of the rare-event enumerations from both staining protocols with matched samples revealed various positive and negative

detected by both IF protocols were positively correlated, such as the DAPI-only cells ($p = 0.032$, $\tau = 0.51$) and the CK-only cells ($p = 0.018$, $\tau = 0.55$). Similar positive correlations were found among the aggregate classifications, including the oncosomes ($p = 0.003$, $\tau = 0.65$) and total CK-positive rare-event population ($p = 0.006$, $\tau = 0.62$). A positive association between rare events that differ by one unshared biomarker between the stains potentially indicates a single rare-event type. An example of this is the CK-only cell counts from the Landscape-staining protocol being positively associated with the CK|CDX2 cells from the CDX2 IF protocol ($p = 0.029$, $\tau = 0.51$). A similar pattern was observed within the oncosome population, with the Onc CK|Vim being positively correlated to the Onc CK|CDX2 ($p < 0.001$, $\tau = 0.82$). The significant negative correlations across the assays included the DAPI-only cells from Landscape-staining protocol and the CK|CD45 cells from the CDX2 IF protocol ($p = 0.029$, $\tau = -0.51$). Interestingly, if selected for rarity, the CD45/CD31 population from the Landscape-staining protocol was uncorrelated with the CD45 population from the CDX2-staining protocol ($p = 0.645$, $\tau = 0.12$), indicating that the differential IF expression was likely due to the addition of the CD31 biomarker.
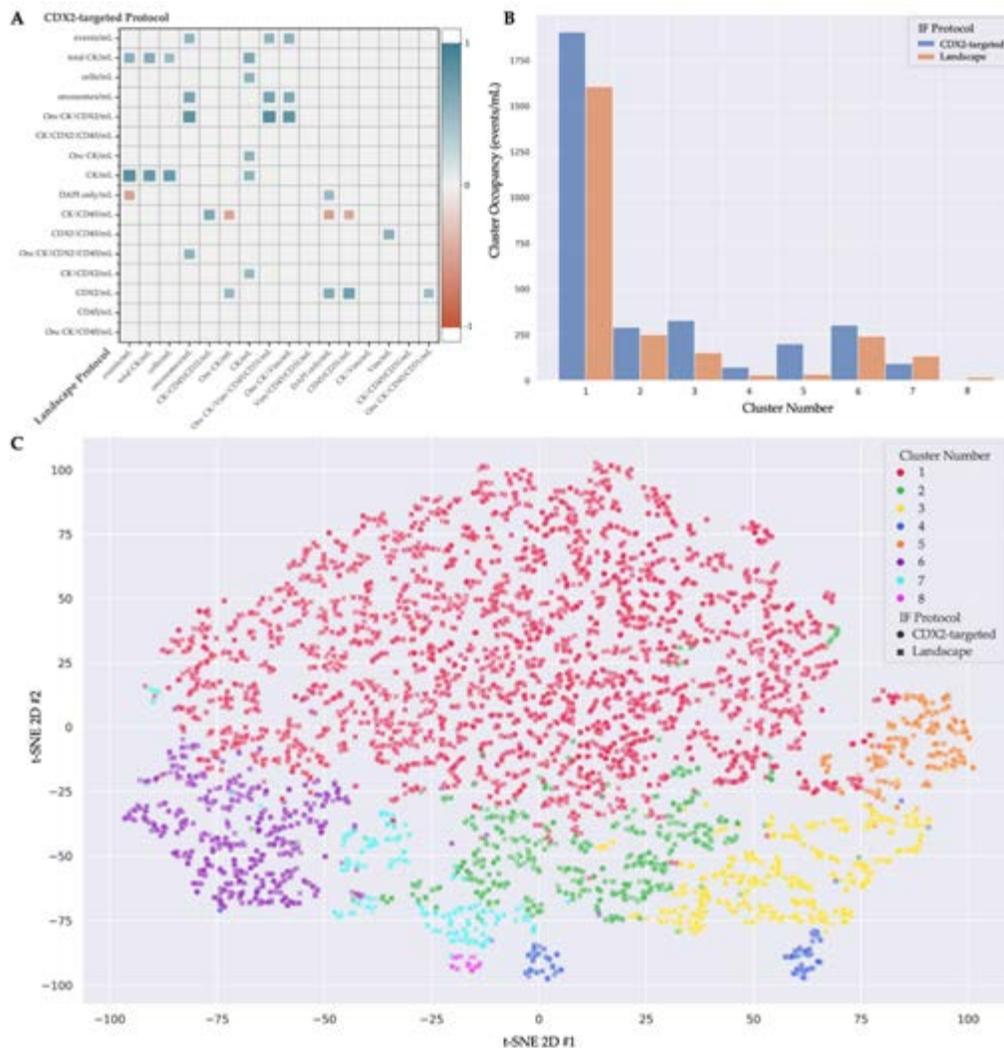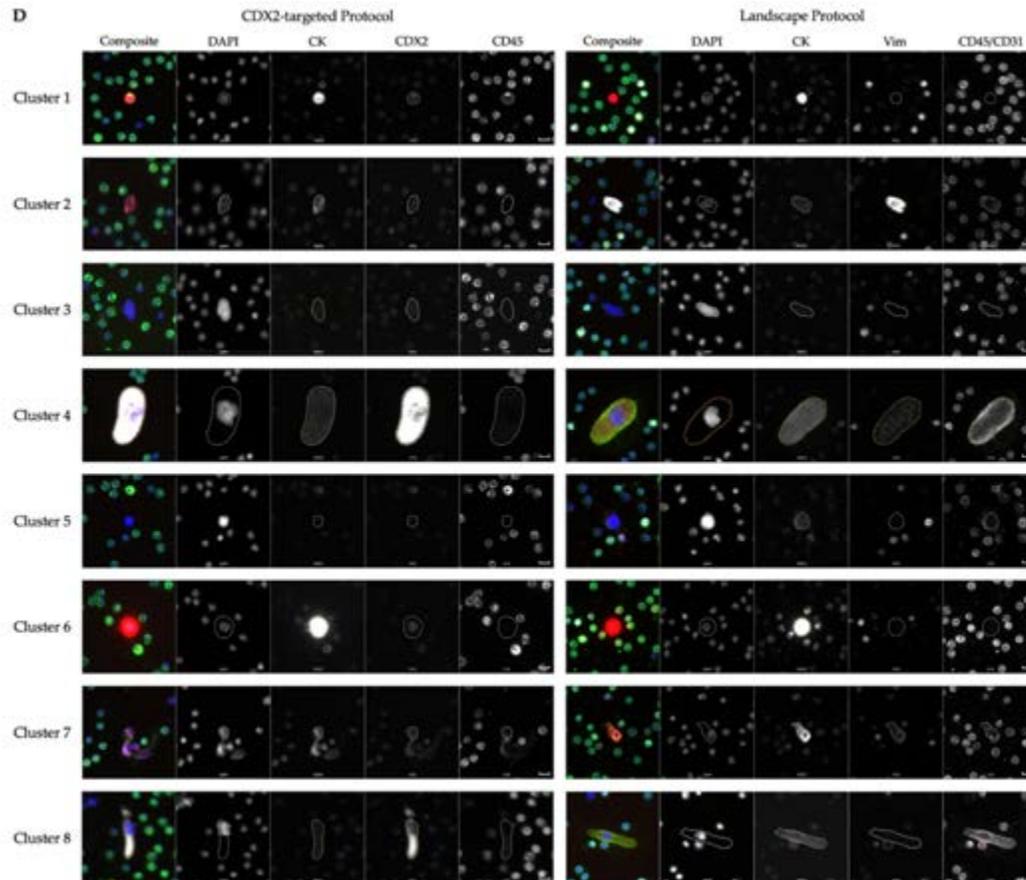

Figure 4. *Cont.*

Figure 4. Multi-assay analysis of the Landscape and CDX2-targeted immunofluorescence (IF) assays analyzed by OCULAR. (**A**) Statistically significant ($p \leq 0.05$) rare-event-count correlations across the two IF assays, with the red-to-blue color gradient indicating a negative-to-positive correlation, respectively. (**B**) Cluster occupancy of the rare cells identified by the CDX2-targeted and Landscape stains when using the 8-group hierarchical clustering model. (**C**) t-Stochastic neighbor embedding (t-SNE) plot of the 8 cell clusters comprised of rare cells from both stains, as indicated by the markers' shape. (**D**) Representative gallery of the 8 clusters with cells from both assays. Each row represents a cluster with a cell from the CDX2-targeted protocol on the left and a cell from the Landscape protocol on the right. Images taken at 100× magnification. Scale bars represent 100 μm.

OCULAR presents morphometrics related to the size and shape of the cell and nucleus. A clustering analysis of the morphometric features was conducted to characterize the rare-event types across the two IF staining protocols, improving our understanding of the analytes in liquid biopsy. Hierarchical clustering into eight groups afforded the most discrete separation of the hypothesized megakaryocytes and endothelial cells. It is important to note that these highlighted rare cell types are not entirely separated into their own clusters, indicating the morphological heterogeneity within the cell categories. The distribution of cells from both IF staining protocols into the eight clusters and representative images are provided in Figure 4B–D. Table 2 provides a description of each cluster.

Table 2. Cellular description for the multi-assay cluster analysis. Cluster occupancy of the rare cells identified by the CDX2-targeted and Landscape staining protocols when using the 8-group hierarchical clustering model for multi-assay analysis.

Cluster Number	Cluster Comments	Cells from Landscape	Comments from Landscape	Cells from CDX2-Targeted	Comments from CDX2-Targeted
1	Heterogeneous phenotype with cellular morphology similar to WBCs	1605	Most prominent: CK|Vim|CD45/CD31	1903	Most prominent: CK|CDX2|CD45
2	Includes endothelial cells	249	Morphologically distinct Vim|CD45/CD31 cells with variable CK expression	291	64 (22%) DAPI-only, 67 (23%) CK and 34 (12%) CK|CDX2 cells
3	Large nuclei, more eccentric than cluster 5	152	101 (66%) DAPI only, 11 (7%) morphologically distinct CD45/CD31 expressing cells	328	279 (85%) DAPI-only
4	Includes megakaryocytes	30	25 (83%) morphologically distinct CD45-/CD31-expressing cells, 5 (17%) small rod-like CD45-/CD31-expressing cells	72	DAPI-only and CDX2-only with similar large morphology
5	Large nuclei, more circular than cluster 3	34	15 (44%) DAPI-only	201	191 (95%) DAPI-only
6	CK only CTCs	242	219 (90%) Epi.CTCs	302	265 (88%) CK and 29 (10%) CK|CDX2 cells
7	Includes endothelial cells	134	84 (63%) morphologically distinct Vim|CD45/CD31 cells with variable CK expression	93	80 (86%) CK only. Morphologically distinct from cluster 6, more elongated.
8	Includes megakaryocytes	18	Morphologically distinct CD45/CD31 cells with variable CK and Vim expression	6	Large, morphologically distinct cells with punctate CDX2 expression

3.4. Clinical Correlation of Liquid-Biopsy Data

Next, we investigated the clinical relevance of the rare events detected in the liquid biopsy. PFS was reported for nine of the ten patients (mean PFS = 6.98 months). A survival analysis of the rare events identified by the Landscape assay in Draw 1 samples revealed a positive correlation between the number of Onc CK\|CD45/CD31 and PFS ($p = 0.0372$, $\tau = 0.70$), as patients with ≥ 2.21 oncosomes/mL had an improved survival (Figure 5A). A similar analysis of the populations identified by the CDX2-targeted assay in Draw 2 samples revealed two rare-event types negatively associated with PFS: CK\|CD45 cells ($p = 0.0362$, $\tau = -0.77$) and Onc CK ($p = 0.0068$, $\tau = -0.89$). Patients with ≥ 20.61 CK\|CD45 cells/mL or ≥ 11.57 Onc CK/mL had poor survival (Figure 5B).

Beyond timepoint-static enumerations, the rare-event kinetics (Figure 5C,D), defined as the change in rare-event subtypes between draws, were significantly associated with PFS. Of the nine patients with survival data, two were missing Draw 2; therefore, only seven patients were included in the kinetics survival analysis. The changes in the CK\|CDX2\|CD45 cells ($p = 0.0068$, $\tau = 0.89$) and the Onc CK\|Vim\|CD45/CD31 ($p = 0.0025$, $\tau = 0.93$) were
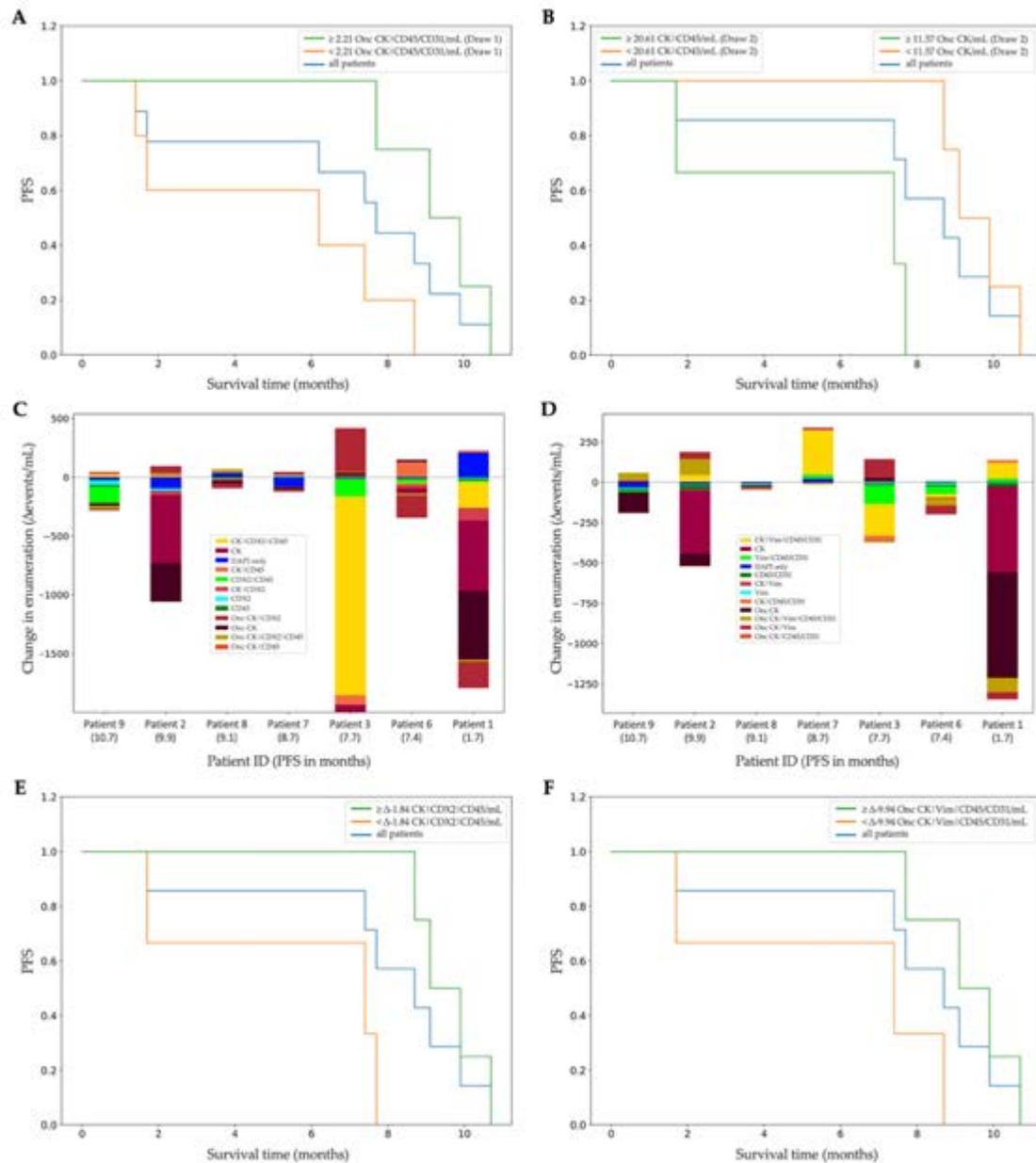
Figure 5. Survival analysis of rare events from the Landscape- and CDX2-targeted immunofluorescence protocols. (**A**) Kaplan-Meier (KM) curve showing that patients with more than the median of 2.21 Onc CK|CD45/CD31/mL found in Draw 1 by the Landscape protocol had longer progression-free survival (PFS). (**B**) KM curve showing two rare events with survival implications found in Draw 2 by the CDX2-targeted protocol. Patients with more than the median of 20.61 CK|CD45 cells/mL or more than the median of 11.57 Onc CK/mL had shorter PFS. (**C**) Rare-event kinetics between Draw 1 and Draw 2 analyzed with the CDX2-targeted protocol, ordered from longest to shortest PFS. (**D**) Rare-event kinetics between Draw 1 and Draw 2 analyzed with the Landscape protocol, ordered from longest to shortest PF€(**E**) KM curve showing that patients with a change of ≥−1.84 CK|CDX2|CD45 cells/mL from the CDX2-targeted protocol had longer PFS. (**F**) KM curve

4. Discussion

As a minimally invasive approach, the liquid biopsy has the potential to significantly advance patient care by addressing current clinical challenges in mCRC. In this study, we show a comprehensive profile of the liquid biopsy that encapsulates the heterogeneous CTC and oncosome populations while providing the initial clinical validation of the liquid biopsy in mCRC patients.

This study corroborates findings of a limited CK-only CTC population in mCRC [25–27]. mCRC is known for its extensive solid tumor heterogeneity, and herein, we show interpatient and intrapatient heterogeneity in the circulating rare-event population. The Landscape protocol aided in the phenotypic characterization of morphologically distinct cell types, allowing for the detection of new rare-cell populations, such as circulating endothelial cells (CECs). Initially identified in the 1970s [60], CEC counts have been shown to be elevated in the liquid biopsy of cancer patients [61]. In mCRC, tumor-derived CECs have been used as a prognostic indicator of clinical response to first-line therapies [62] and patient survival [41,42]. Interestingly, a previous investigation targeting CTCs in CRC instead discovered tumor-derived CEC clusters that starkly differentiated normal donors, treatment-naïve and early-stage patients [63]. The detection of clinically significant CECs while attempting to find CTCs highlights the importance of using an unbiased rare-cell approach to capturing the heterogeneity in the liquid biopsy of mCRC. In this study, the morphologically distinct Vim | CD45/CD31 cells with variable CK expression were found individually and in clusters. Based on their morphology and biomarker expression, we expect that these cells are CECs. While downstream proteomics and genomics are needed to confirm this cell lineage, the filamentous Vim and punctate CD45/CD31 signal captured by the Landscape IF protocol are characteristic of endothelial cells [64,65]. HDSCA has previously identified CECs in the liquid biopsy, the likes of which present a similar morphology and biomarker expression pattern [33].

We additionally identified large, morphologically distinct CD45/CD31 cells in the mCRC samples that appear to be multilobular with sizeable nuclei and punctate cytoplasmic expression. Such cellular features lead us to hypothesize that these events are megakaryocytes. These platelet-producing cells derived from the bone marrow are positive in CD31, have a granular cytoplasm and are large, with up to a 160 µm diameter [66]. Found either in the solid tumor microenvironment or in circulation, megakaryocytes have shown to have prognostic potential in prostate [67] and non-small-cell lung cancer [68]. While platelet indices have served as diagnostic biomarkers in mCRC [69], the novel identification of potential megakaryocytes in this cancer type could have significant clinical implications even in the absence of direct platelet detection. Future studies should characterize these megakaryocyte candidates with downstream proteomics to confirm their lineage and to compare their enumerations to patient-specific clinical factors and survival. This unique cell type further exemplifies the heterogeneity of the circulating cell profile of mCRC and highlights the power of rare-event detection systems to represent it.

This study demonstrates that the utility of the liquid biopsy is not limited to cellular events, emphasizing the importance of detecting vesicles. Vesicles have been described and classified according to the mechanism of cellular release and size, which may be dependent on the method of detection or isolation [44,70]. Oncosomes comprised a significant portion of the rare-event profiles from both the CDX2-targeted and Landscape assays in mCRC patient samples. HDSCA has previously identified oncosomes with similar size and biomarker expression in prostate cancer [46], bladder cancer [48] and upper tract urothelial carcinoma [51], suggesting that these events may be found in a variety of cancer types. Most importantly, oncosome enumerations from both IF protocols correlated with PFS in this cohort. The kinetics of a Landscape-stained oncosome population also correlated significantly with PFS, pointing to the importance of enumerating these analytes over time. The potential for making diagnoses, prognoses and subsequent treatment decisions based off oncosomes, especially in a cancer type like mCRC that does not widely

Herein, we show that oncosomes associated with mCRC tumorigenesis may be useful prognostic biomarkers.

Predicated on this study's sample-matched design, the multi-assay analysis is a novel attempt to overlap IF protocols using the shared features of an unbiased rare-event detection platform. The successful grouping of known cell types into delineated clusters, such as the Epi.CTCs in cluster 6, serves as a proof of concept for applications in a larger cohort. Using only image-based morphometrics, the multi-assay analysis increases the number of evaluable IF biomarkers for a single sample, while maintaining a low monetary cost relative to single-cell proteomics. Cluster 4 best depicts the value of additional biomarkers in understanding the cellular biology. Using the Landscape assay, we detected megakaryocytes with a CD45/CD31-positive signal, while in the CDX2-targeted assay, we detected megakaryocyte-like cells that presented a CDX2-positive but CD45-negative signal. The shared feature analysis allowed for the identification of these cells as similar and the observation of CDX2 antibody binding to these cells. A multi-assay approach is uniquely suited for rare-event analysis, wherein thresholds from multiple IF protocols conducted in tandem could demonstrate the clinical utility of the liquid biopsy.

Phenotypic switches are fundamental to CRC initiation, metastasis and relapse [71–73], thus requiring longitudinal prognostic tools and changing therapies to target the continuously evolving cell types. The minimally invasive liquid biopsy has the potential to address this challenge, and we show initial evidence for clinical utility. The molecular characterization of the rare events detected in this study will elucidate their potential role in mCRC tumorigenesis. The HDSCA3.0 workflow includes the capability for genomic analysis, both SNV (single-nucleotide variation) and CNV (copy-number variation), for both single cells and cell-free DNA (cfDNA) [74–76], as well as targeted multiplexed proteomic analysis [38,75] on samples previously characterized at the morphological and phenotypic level by IF. The data presented here serves to motivate further genomic and/or proteomic analysis of the CTCs and oncosomes detected to validate their neoplastic origin and association with the disease state.

Additional studies are needed with a greater patient sample size, PB draws from multiple timepoints throughout treatment and patient-specific clinical information (KRAS/NRAS status, TNM staging, etc.) to provide the prognostic, diagnostic and predictive utility of the liquid biopsy in the management of mCRC.

5. Conclusions

As one of the world's most prevalent oncological diseases, mCRC poses numerous clinical challenges due to its extensive tumor heterogeneity. This study establishes evidence for the clinical validation of an unbiased rare-event approach to the liquid biopsy. For the first time, we demonstrate the value of a comprehensive CTC and oncosome detection approach in the PB of mCRC. Our results highlight the heterogeneity of the liquid-biopsy profile with the identification of rare-event frequencies unique to mCRC patients. By analyzing both IF expression patterns and morphological parameters, we identify two select rare cell types that warrant future study into their implications in mCRC. Furthermore, we demonstrate the utility of analyzing multiple IF assays in tandem to characterize the heterogeneous populations detected. These findings motivate the further molecular characterization of these analytes and investigation into their predictive power with respect to patient outcomes in mCRC.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cancers14194891/s1, Figure S1. Enumeration of the rare events detected in the Landscape-stained ND samples analyzed by OCULAR. Table S1. Rare-event frequencies, enumerations, and sample positivity from the enrichment-based and OCULAR analysis of the samples stained with the CDX2-targeted assay. The sample positivity threshold of ≥5 events/mL was determined by comparisons to the rare-event enumerations of an ND cohort. Within each detection approach, the frequency of each classification is provided as a percentage of the total rare-event profile. Figure S2. CK-positive rare events from the CDX2-stained samples analyzed by OCULAR.

(A) Cells positioned on the edge of frames that are intentionally removed by the quality-control process in OCULAR, representative of the 2.04% non-concordance group between the two detection approaches. Frame edges are indicated by the dashed white lines. (B) CK-positive cells, containing representative HD-CTC, CTC-Apoptotic and two CTC-Small candidates as per the CK-focused approach. (C) CK|CDX2 cells, containing a CTC cluster, cells with a dim CK signal, and eccentric rare cells. (D) CK|CD45 cells with varying CD45 expression. (E) CK|CDX2|CD45 cells. (F) Heterogenous CK positive oncosomes. DAPI: blue, CK: red, CDX2: white, CD45: green. Images taken at 100× magnification. Scale bars represent 10 μm. Figure S3. CDX2-stained samples rare-event enumeration using HDSCA1.0. (A) Enumeration (CTCs/ml) of each CTC subtype using the CTC-focused approach per patient and draw. (B) Frequency (%) of each CTC subtype using the CTC-focused approach per patient and draw.

References

1. Siegel, R.L.; Miller, K.D.; Goding Sauer, A.; Fedewa, S.A.; Butterly, L.F.; Anderson, J.C.; Cercek, A.; Smith, R.A.; Jemal, A. Colorectal cancer statistics, 2020. *CA A Cancer J. Clin.* **2020**, *70*, 145–164. [CrossRef] [PubMed]
2. Barker, N.; Bartfeld, S.; Clevers, H. Tissue-resident adult stem cell populations of rapidly self-renewing organs. *Cell Stem Cell* **2010**, *7*, 656–670. [CrossRef] [PubMed]
3. Fontana, E.; Nyamundanda, G.; Cunningham, D.; Tu, D.; Cheang, M.C.; Jonker, D.J.; Siu, L.L.; Sclafani, F.; Eason, K.; Ragulan, C. Intratumoral transcriptome heterogeneity is associated with patient prognosis and sidedness in patients with colorectal cancer treated with anti-EGFR therapy from the CO. 20 trial. *JCO Precis. Oncol.* **2020**, *4*, 1152–1162. [CrossRef]
4. Suzuki, Y.; Ng, S.B.; Chua, C.; Leow, W.Q.; Chng, J.; Liu, S.Y.; Ramnarayanan, K.; Gan, A.; Ho, D.L.; Ten, R. Multiregion ultra-deep sequencing reveals early intermixing and variable levels of intratumoral heterogeneity in colorectal cancer. *Mol. Oncol.* **2017**, *11*, 124–139. [CrossRef]
5. Moorcraft, S.Y.; Smyth, E.C.; Cunningham, D. The role of personalized medicine in metastatic colorectal cancer: An evolving landscape. *Ther. Adv. Gastroenterol.* **2013**, *6*, 381–395. [CrossRef] [PubMed]
6. Molinari, C.; Marisi, G.; Passardi, A.; Matteucci, L.; De Maio, G.; Ulivi, P. Heterogeneity in colorectal cancer: A challenge for personalized medicine? *Int. J. Mol. Sci.* **2018**, *19*, 3733. [CrossRef]
7. Kyrochristos, I.D.; Roukos, D.H. Comprehensive intra-individual genomic and transcriptional heterogeneity: Evidence-based Colorectal Cancer Precision Medicine. *Cancer Treat. Rev.* **2019**, *80*, 101894. [CrossRef] [PubMed]
8. Chen, D.; Xu, T.; Wang, S.; Chang, H.; Yu, T.; Zhu, Y.; Chen, J. Liquid Biopsy Applications in the Clinic. *Mol. Diagn. Ther.* **2020**, *24*, 125–132.

9.   Esposito, A.; Criscitiello, C.; Locatelli, M.; Milano, M.; Curigliano, G. Liquid biopsies for solid tumors: Understanding tumor heterogeneity and real time monitoring of early resistance to targeted therapies. *Pharmacol. Ther.* **2016**, *157*, 120–124. [CrossRef]

10. Russano, M.; Napolitano, A.; Ribelli, G.; Iuliani, M.; Simonetti, S.; Citarella, F.; Pantano, F.; Dell'Aquila, E.; Anesi, C.; Silvestris, N. Liquid biopsy and tumor heterogeneity in metastatic solid tumors: The potentiality of blood samples. *J. Exp. Clin. Cancer Res.* **2020**, *39*, 95. [CrossRef]

11. Van Cutsem, E.; Cervantes, A.; Nordlinger, B.; Arnold, D. Metastatic colorectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **2014**, *25*, iii1–iii9. [CrossRef] [PubMed]

12. Benson, A.B.; Venook, A.P.; Al-Hawary, M.M.; Arain, M.A.; Chen, Y.-J.; Ciombor, K.K.; Cohen, S.; Cooper, H.S.; Deming, D.; Farkas, L.; et al. Colon Cancer, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Cancer Netw.* **2021**, *19*, 329–359. [CrossRef] [PubMed]

13. Aggarwal, C.; Meropol, N.; Punt, C.; Iannotti, N.; Saidman, B.; Sabbath, K.; Gabrail, N.; Picus, J.; Morse, M.; Mitchell, E. Relationship among circulating tumor cells, CEA and overall survival in patients with metastatic colorectal cancer. *Ann. Oncol.* **2013**, *24*, 420–428. [CrossRef] [PubMed]

14. Cohen, S.J.; Punt, C.J.; Iannotti, N.; Saidman, B.H.; Sabbath, K.D.; Gabrail, N.Y.; Picus, J.; Morse, M.; Mitchell, E.; Miller, M.C.; et al. Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. *J. Clin. Oncol.* **2008**, *26*, 3213–3221. [CrossRef] [PubMed]

15. Hendricks, A.; Brandt, B.; Geisen, R.; Dall, K.; Röder, C.; Schafmayer, C.; Becker, T.; Hinz, S.; Sebens, S. Isolation and enumeration of CTC in colorectal cancer patients: Introduction of a novel cell imaging approach and comparison to cellular and molecular detection techniques. *Cancers* **2020**, *12*, 2643. [CrossRef] [PubMed]

16. Kolenčík, D.; Narayan, S.; Thiele, J.-A.; McKinley, D.; Gerdtsson, A.S.; Welter, L.; Hošek, P.; Ostašov, P.; Vyčítal, O.; Brůha, J.; et al. Circulating tumor cell kinetics and morphology from the liquid biopsy predict disease progression in patients with metastatic colorectal cancer following resection. *Cancers* **2022**, *14*, 642. [CrossRef] [PubMed]

17. Kolencik, D.; Shishido, S.N.; Pitule, P.; Mason, J.; Hicks, J.; Kuhn, P. Liquid biopsy in colorectal carcinoma: Clinical applications and challenges. *Cancers* **2020**, *12*, 1376. [CrossRef]

18. Veridex LLC. CellSearch Circulating Tumor Cell Kit Premarket Notification-Expanded Indications for Use—Metastatic Prostate Cancer. 2008. Available online: https://www.accessdata.fda.gov/cdrh_docs/pdf7/k073338.pdf (accessed on 16 December 2008).

19. Allard, W.J.; Matera, J.; Miller, M.C.; Repollet, M.; Connelly, M.C.; Rao, C.; Tibbe, A.G.; Uhr, J.W.; Terstappen, L.W. Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clin. Cancer Res.* **2004**, *10*, 6897–6904. [CrossRef]

20. Miller, M.C.; Doyle, G.V.; Terstappen, L.W. Significance of circulating tumor cells detected by the CellSearch system in patients with metastatic breast colorectal and prostate cancer. *J. Oncol.* **2010**, *2010*, 617421. [CrossRef]

21. Seeberg, L.T.; Waage, A.; Brunborg, C.; Hugenschmidt, H.; Renolen, A.; Stav, I.; Bjornbeth, B.A.; Brudvik, K.W.; Borgen, E.F.; Naume, B.; et al. Circulating tumor cells in patients with colorectal liver metastasis predict impaired survival. *Ann. Surg.* **2015**, *261*, 164–171. [CrossRef]

22. Huang, X.; Gao, P.; Song, Y.; Sun, J.; Chen, X.; Zhao, J.; Xu, H.; Wang, Z. Meta-analysis of the prognostic value of circulating tumor cells detected with the CellSearch System in colorectal cancer. *BMC Cancer* **2015**, *15*, 202. [CrossRef] [PubMed]

23. Mazouji, O.; Ouhajjou, A.; Incitti, R.; Mansour, H. Updates on clinical use of liquid biopsy in colorectal cancer screening, diagnosis, follow-up, and treatment guidance. *Front. Cell Dev. Biol.* **2021**, *9*, 962. [CrossRef] [PubMed]

24. Arneth, B. Update on the types and usage of liquid biopsies in the clinical setting: A systematic review. *BMC Cancer* **2018**, *18*, 527. [CrossRef]

25. Gerdtsson, A.S.; Thiele, J.A.; Shishido, S.N.; Zheng, S.; Schaffer, R.; Bethel, K.; Curley, S.; Lenz, H.J.; Hanna, D.L.; Nieva, J.; et al. Single cell correlation analysis of liquid and solid biopsies in metastatic colorectal cancer. *Oncotarget* **2019**, *10*, 7016–7030. [CrossRef] [PubMed]

26. Sotelo, M.; Sastre, J.; Maestro, M.; Veganzones, S.; Vieitez, J.; Alonso, V.; Gravalos, C.; Escudero, P.; Vera, R.; Aranda, E. Role of circulating tumor cells as prognostic marker in resected stage III colorectal cancer. *Ann. Oncol.* **2015**, *26*, 535–541. [CrossRef]

27. Wang, L.; Balasubramanian, P.; Chen, A.P.; Kummar, S.; Evrard, Y.A.; Kinders, R.J. Promise and limits of the CellSearch platform for evaluating pharmacodynamics in circulating tumor cells. *Semin. Oncol.* **2016**, *43*, 464–475. [CrossRef] [PubMed]

28. Shishido, S.N.; Carlsson, A.; Nieva, J.; Bethel, K.; Hicks, J.B.; Bazhenova, L.; Kuhn, P. Circulating tumor cells as a response monitor in stage IV non-small cell lung cancer. *J. Transl. Med.* **2019**, *17*, 294. [CrossRef]

29. Keomanee-Dizon, K.; Shishido, S.N.; Kuhn, P. Circulating tumor cells: High-throughput imaging of CTCs and bioinformatic analysis. *Recent Results Cancer Res.* **2020**, *215*, 89–104. [CrossRef]

30. Marrinucci, D.; Bethel, K.; Kolatkar, A.; Luttgen, M.S.; Malchiodi, M.; Baehring, F.; Voigt, K.; Lazar, D.; Nieva, J.; Bazhenova, L. Fluid biopsy in patients with metastatic prostate, pancreatic and breast cancers. *Phys. Biol.* **2012**, *9*, 016003. [CrossRef]

31. Rodríguez-Lee, M.; Kolatkar, A.; McCormick, M.; Dago, A.D.; Kendall, J.; Carlsson, N.A.; Bethel, K.; Greenspan, E.J.; Hwang, S.E.; Waitman, K.R. Effect of blood collection tube type and time to processing on the enumeration and high-content characterization of circulating tumor cells using the high-definition single-cell assay. *Arch. Pathol. Lab. Med.* **2018**, *142*, 198–207. [CrossRef]

32. Shishido, S.N.; Welter, L.; Rodriguez-Lee, M.; Kolatkar, A.; Xu, L.; Ruiz, C.; Gerdtsson, A.S.; Restrepo-Vassalli, S.; Carlsson, A.; Larsen, J. Preanalytical variables for the genomic assessment of the cellular and acellular fractions of the liquid biopsy in a cohort of breast cancer patients. *J. Mol. Diagn.* **2020**, *22*, 319–337. [CrossRef] [PubMed]

33. Bethel, K.; Luttgen, M.S.; Damani, S.; Kolatkar, A.; Lamy, R.; Sabouri-Ghomi, M.; Topol, S.; Topol, E.J.; Kuhn, P. Fluid phase biopsy for detection and characterization of circulating endothelial cells in myocardial infarction. *Phys. Biol.* **2014**, *11*, 016002. [CrossRef] [PubMed]

34. Armstrong, A.J.; Halabi, S.; Luo, J.; Nanus, D.M.; Giannakakou, P.; Szmulewitz, R.Z.; Danila, D.C.; Healy, P.; Anand, M.; Rothwell, C.J. Prospective multicenter validation of androgen receptor splice variant 7 and hormone therapy resistance in high-risk castration-resistant prostate cancer: The prophecy study. *J. Clin. Oncol.* **2019**, *37*, 1120. [CrossRef] [PubMed]

35. Scher, H.I.; Graf, R.P.; Schreiber, N.A.; Jayaram, A.; Winquist, E.; McLaughlin, B.; Lu, D.; Fleisher, M.; Orr, S.; Lowes, L. Assessment of the validity of nuclear-localized androgen receptor splice variant 7 in circulating tumor cells as a predictive biomarker for castration-resistant prostate cancer. *JAMA Oncol.* **2018**, *4*, 1179–1186. [CrossRef] [PubMed]

36. Scher, H.I.; Lu, D.; Schreiber, N.A.; Louw, J.; Graf, R.P.; Vargas, H.A.; Johnson, A.; Jendrisak, A.; Bambury, R.; Danila, D. Association of AR-V7 on circulating tumor cells as a treatment-specific biomarker with outcomes and survival in castration-resistant prostate cancer. *JAMA Oncol.* **2016**, *2*, 1441–1449. [CrossRef]

37. Welter, L.; Xu, L.; McKinley, D.; Dago, A.E.; Prabakar, R.K.; Restrepo-Vassalli, S.; Xu, K.; Rodriguez-Lee, M.; Kolatkar, A.; Nevarez, R. Treatment response and tumor evolution: Lessons from an extended series of multianalyte liquid biopsies in a metastatic breast cancer patient. *Mol. Case Stud.* **2020**, *6*, a005819. [CrossRef]

38. Gerdtsson, E.; Pore, M.; Thiele, J.-A.; Gerdtsson, A.S.; Malihi, P.D.; Nevarez, R.; Kolatkar, A.; Velasco, C.R.; Wix, S.; Singh, M. Multiplex protein detection on circulating tumor cells from liquid biopsies using imaging mass cytometry. *Converg. Sci. Phys. Oncol.* **2018**, *4*, 015002. [CrossRef]

39. Febbo, P.G.; Martin, A.M.; Scher, H.I.; Barrett, J.C.; Beaver, J.A.; Beresford, P.J.; Blumenthal, G.M.; Bramlett, K.; Compton, C.; Dittamore, R. Minimum technical data elements for liquid biopsy data submitted to public databases. *Clin. Pharmacol. Ther.* **2020**, *107*, 730–734. [CrossRef]

40. Ghatalia, P.; Smith, C.H.; Winer, A.; Gou, J.; Kiedrowski, L.A.; Slifker, M.; Saltzberg, P.D.; Bubes, N.; Anari, F.M.; Kasireddy, V.; et al. Clinical Utilization Pattern of Liquid Biopsies (LB) to Detect Actionable Driver Mutations, Guide Treatment Decisions and Monitor Disease Burden During Treatment of 33 Metastatic Colorectal Cancer (mCRC) Patients (pts) at a Fox Chase Cancer Center GI Oncology Subspecialty Clinic. *Front. Oncol.* **2018**, *8*, 652. [CrossRef]

41. Matsusaka, S.; Suenaga, M.; Mishima, Y.; Takagi, K.; Terui, Y.; Mizunuma, N.; Hatake, K. Circulating endothelial cells predict for response to bevacizumab-based chemotherapy in metastatic colorectal cancer. *Cancer Chemother. Pharmacol.* **2011**, *68*, 763–768. [CrossRef]

42. Simkens, L.; Tol, J.; Terstappen, L.; Teerenstra, S.; Punt, C.; Nagtegaal, I. The predictive and prognostic value of circulating endothelial cells in advanced colorectal cancer patients receiving first-line chemotherapy and bevacizumab. *Ann. Oncol.* **2010**, *21*, 2447–2448. [CrossRef] [PubMed]

43. Okusha, Y.; Eguchi, T.; Tran, M.T.; Sogawa, C.; Yoshida, K.; Itagaki, M.; Taha, E.A.; Ono, K.; Aoyama, E.; Okamura, H. Extracellular vesicles enriched with moonlighting metalloproteinase are highly transmissive, pro-tumorigenic, and trans-activates cellular communication network factor (CCN2/CTGF): CRISPR against cancer. *Cancers* **2020**, *12*, 881. [CrossRef] [PubMed]

44. Di Vizio, D.; Kim, J.; Hager, M.H.; Morello, M.; Yang, W.; Lafargue, C.J.; True, L.D.; Rubin, M.A.; Adam, R.M.; Beroukhim, R. Oncosome formation in prostate cancer: Association with a region of frequent chromosomal deletion in metastatic disease. *Cancer Res.* **2009**, *69*, 5601–5609. [CrossRef] [PubMed]

45. Ciardiello, C.; Leone, A.; Lanuti, P.; Roca, M.S.; Moccia, T.; Minciacchi, V.R.; Minopoli, M.; Gigantino, V.; De Cecio, R.; Rippa, M. Large oncosomes overexpressing integrin alpha-V promote prostate cancer adhesion and invasion via AKT activation. *J. Exp. Clin. Cancer Res.* **2019**, *38*, 317. [CrossRef]

46. Gerdtsson, A.S.; Setayesh, S.M.; Malihi, P.D.; Ruiz, C.; Carlsson, A.; Nevarez, R.; Matsumoto, N.; Gerdtsson, E.; Zurita, A.; Logothetis, C.; et al. Large extracellular vesicle characterization and association with circulating tumor cells in metastatic castrate resistant prostate cancer. *Cancers* **2021**, *13*, 1056. [CrossRef]

47. Chai, S.; Matsumoto, N.; Storgard, R.; Peng, C.C.; Aparicio, A.; Ormseth, B.; Rappard, K.; Cunningham, K.; Kolatkar, A.; Nevarez, R.; et al. Platelet-coated circulating tumor cells are a predictive biomarker in patients with metastatic castrate-resistant prostate cancer. *Mol. Cancer Res.* **2021**, *19*, 2036–2045. [CrossRef]

48. Shishido, S.N.; Sayeed, S.; Courcoubetis, G.; Djaladat, H.; Miranda, G.; Pienta, K.J.; Nieva, J.; Hansel, D.E.; Desai, M.; Gill, I.S.; et al. Characterization of cellular and acellular analytes from pre-cystectomy liquid biopsies in patients newly diagnosed with primary bladder cancer. *Cancers* **2022**, *14*, 758. [CrossRef]

49. Werling, R.W.; Yaziji, H.; Bacchi, C.E.; Gown, A.M. CDX2, a highly sensitive and specific marker of adenocarcinomas of intestinal origin: An immunohistochemical survey of 476 primary and metastatic carcinomas. *Am. J. Surg. Pathol.* **2003**, *27*, 303–310. [CrossRef]

50. Adams, D.L.; Stefansson, S.; Haudenschild, C.; Martin, S.S.; Charpentier, M.; Chumsri, S.; Cristofanilli, M.; Tang, C.M.; Alpaugh, R.K. Cytometric characterization of circulating tumor cells captured by microfiltration and their correlation to the cellsearch® CTC test. *Cytom. Part A* **2015**, *87*, 137–144. [CrossRef]

51. Shishido, S.N.; Ghoreifi, A.; Sayeed, S.; Courcoubetis, G.; Huang, A.; Ye, B.; Mrutyunjaya, S.; Gill, I.S.; Kuhn, P.; Mason, J.; et al. Liquid biopsy landscape in patients with primary upper tract Urothelial Carcinoma. *Cancers* **2022**, *14*, 3007. [CrossRef]

52. Minciacchi, V.R.; Freeman, M.R.; Di Vizio, D. Extracellular vesicles in cancer: Exosomes, microvesicles and the emerging role of large oncosomes. *Semin. Cell Dev. Biol.* **2015**, *40*, 41–51. [CrossRef] [PubMed]

53. Di Vizio, D.; Morello, M.; Dudley, A.C.; Schow, P.W.; Adam, R.M.; Morley, S.; Mulholland, D.; Rotinen, M.; Hager, M.H.; Insabato, L. Large oncosomes in human prostate cancer tissues and in the circulation of mice with metastatic disease. *Am. J. Pathol.* **2012**, *181*, 1573–1584. [CrossRef] [PubMed]
54. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
55. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **1987**, *100*, 441–471. [CrossRef]
56. Kaplan, E.L.; Meier, P. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **1958**, *53*, 457–481. [CrossRef]
57. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef]
58. Pölsterl, S. Scikit-survival: A library for time-to-event analysis built on top of scikit-learn. *J. Mach. Learn. Res.* **2020**, *21*, 8747–8752.
59. Sharpe, A.; McIntosh, M.; Lawrie, S.M. Research methods, statistics and evidence-based practice. In *Companion to Psychiatric Studies*, 8th ed.; Churchill Livingstone: London, UK, 2010; p. 157.
60. Hladovec, J.; Rossmann, P. Circulating endothelial cells isolated together with platelets and the experimental modification of their counts in rats. *Thromb. Res.* **1973**, *3*, 665–674. [CrossRef]
61. Mancuso, P.; Burlini, A.; Pruneri, G.; Goldhirsch, A.; Martinelli, G.; Bertolini, F. Resting and activated endothelial cells are increased in the peripheral blood of cancer patients. *Blood J. Am. Soc. Hematol.* **2001**, *97*, 3658–3661. [CrossRef]
62. Ronzoni, M.; Manzoni, M.; Mariucci, S.; Loupakis, F.; Brugnatelli, S.; Bencardino, K.; Rovati, B.; Tinelli, C.; Falcone, A.; Villa, E. Circulating endothelial cells and endothelial progenitors as predictive markers of clinical response to bevacizumab-based first-line treatment in advanced colorectal cancer patients. *Ann. Oncol.* **2010**, *21*, 2382–2389. [CrossRef]
63. Cima, I.; Kong, S.L.; Sengupta, D.; Tan, I.B.; Phyo, W.M.; Lee, D.; Hu, M.; Iliescu, C.; Alexander, I.; Goh, W.L. Tumor-derived circulating endothelial cell clusters in colorectal cancer. *Sci. Transl. Med.* **2016**, *8*, 345ra389. [CrossRef] [PubMed]
64. Dave, J.M.; Bayless, K.J. Vimentin as an integral regulator of cell adhesion and endothelial sprouting. *Microcirculation* **2014**, *21*, 333–344. [CrossRef] [PubMed]
65. DeLisser, H.; Newman, P.; Albelda, S. Platelet endothelial cell adhesion molecule (CD31). *Curr. Top. Microbiol. Immunol.* **1993**, *184*, 37–45. [PubMed]
66. Metcalf, D.; MacDonald, H.; Odartchenko, N.; Sordat, B. Growth of mouse megakaryocyte colonies in vitro. *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 1744–1748. [CrossRef]
67. Xu, L.; Mao, X.; Guo, T.; Chan, P.Y.; Shaw, G.; Hines, J.; Stankiewicz, E.; Wang, Y.; Oliver, R.T.D.; Ahmad, A.S. The novel association of circulating tumor cells and circulating megakaryocytes with prostate cancer prognosis. *Clin. Cancer Res.* **2017**, *23*, 5112–5122. [CrossRef] [PubMed]
68. Huang, W.; Zhao, S.; Xu, W.; Zhang, Z.; Ding, X.; He, J.; Liang, W. Presence of intra-tumoral CD61+ megakaryocytes predicts poor prognosis in non-small cell lung cancer. *Transl. Lung Cancer Res.* **2019**, *8*, 323. [CrossRef]
69. Zhu, X.; Cao, Y.; Lu, P.; Kang, Y.; Lin, Z.; Hao, T.; Song, Y. Evaluation of platelet indices as diagnostic biomarkers for colorectal cancer. *Sci. Rep.* **2018**, *8*, 11814. [CrossRef] [PubMed]
70. Crescitelli, R.; Lässer, C.; Lötvall, J. Isolation and characterization of extracellular vesicle subpopulations from tissues. *Nature Protocols* **2021**, *16*, 1548–1580. [CrossRef]
71. Schwitalla, S.; Fingerle, A.A.; Cammareri, P.; Nebelsiek, T.; Göktuna, S.I.; Ziegler, P.K.; Canli, O.; Heijmans, J.; Huels, D.J.; Moreaux, G. Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties. *Cell* **2013**, *152*, 25–38. [CrossRef]
72. Merlos-Suárez, A.; Barriga, F.M.; Jung, P.; Iglesias, M.; Céspedes, M.V.; Rossell, D.; Sevillano, M.; Hernando-Momblona, X.; da Silva-Diz, V.; Muñoz, P. The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. *Cell Stem Cell* **2011**, *8*, 511–524. [CrossRef]
73. Fumagalli, A.; Oost, K.C.; Kester, L.; Morgner, J.; Bornes, L.; Bruens, L.; Spaargaren, L.; Azkanaz, M.; Schelfhorst, T.; Beerling, E. Plasticity of Lgr5-negative cancer cells drives metastasis in colorectal cancer. *Cell Stem Cell* **2020**, *26*, 569–578.e7. [CrossRef] [PubMed]
74. Dago, A.E.; Stepansky, A.; Carlsson, A.; Luttgen, M.; Kendall, J.; Baslan, T.; Kolatkar, A.; Wigler, M.; Bethel, K.; Gross, M.E. Rapid phenotypic and genomic change in response to therapeutic pressure in prostate cancer inferred by high content analysis of single circulating tumor cells. *PLoS ONE* **2014**, *9*, e101777. [CrossRef] [PubMed]
75. Malihi, P.D.; Morikado, M.; Welter, L.; Liu, S.T.; Miller, E.T.; Cadaneanu, R.M.; Knudsen, B.S.; Lewis, M.S.; Carlsson, A.; Velasco, C.R. Clonal diversity revealed by morphoproteomic and copy number profiles of single prostate cancer cells at diagnosis. *Converg. Sci. Phys. Oncol.* **2018**, *4*, 015003. [CrossRef] [PubMed]
76. Ruiz, C.; Li, J.; Luttgen, M.S.; Kolatkar, A.; Kendall, J.T.; Flores, E.; Topp, Z.; Samlowski, W.E.; McClay, E.; Bethel, K. Limited genomic heterogeneity of circulating melanoma cells in advanced stage patients. *Phys. Biol.* **2015**, *12*, 016008. [CrossRef]

PNAS

Physicochemical models of protein–DNA binding with standard and modified base pairs

Tsu-Pei Chiu[a] [ID], Satyanarayan Rao[a,1] [ID], and Remo Rohs[a,b,c,d,2] [ID]

DNA-binding proteins play important roles in various cellular processes, but the mechanisms by which proteins recognize genomic target sites remain incompletely understood. Functional groups at the edges of the base pairs (bp) exposed in the DNA grooves represent physicochemical signatures. As these signatures enable proteins to form specific contacts between protein residues and bp, their study can provide mechanistic insights into protein–DNA binding. Existing experimental methods, such as X-ray crystallography, can reveal such mechanisms based on physicochemical interactions between proteins and their DNA target sites. However, the low throughput of structural biology methods limits mechanistic insights for selection of many genomic sites. High-throughput binding assays enable prediction of potential target sites by determining relative binding affinities of a protein to massive numbers of DNA sequences. Many currently available computational methods are based on the sequence of standard Watson–Crick bp. They assume that the contribution of overall binding affinity is independent for each base pair, or alternatively include dinucleotides or short k-mers. These methods cannot directly expand to physicochemical contacts, and they are not suitable to apply to DNA modifications or non-Watson–Crick bp. These variations include DNA methylation, and synthetic or mismatched bp. The proposed method, DeepRec, can predict relative binding affinities as function of physicochemical signatures and the effect of DNA methylation or other chemical modifications on binding. Sequence-based modeling methods are in comparison a coarse-grain description and cannot achieve such insights. Our chemistry-based modeling framework provides a path towards understanding genome function at a mechanistic level.

transcription factor | binding specificity | readout mode | quantitative modeling | deep learning

DNA-binding proteins selectively bind to their genomic binding sites and regulate various cellular processes. This selective binding occurs when the DNA-binding domain of a gene regulatory protein, a transcription factor (TF), recognizes its binding site by reading physicochemical signatures at the base-pair (bp) edges (Fig. 1*A*). These physicochemical signatures, consisting of hydrogen bond (H-bond) acceptors, H-bond donors, methyl groups, and nonpolar hydrogen atoms, are exposed on the surface of the DNA major and minor grooves (Fig. 1*B*) and allow protein residues to form a series of chemical contacts, including H-bonds, water-mediated H-bonds, and hydrophobic interactions (1).

Structural information from TF–DNA complexes provides insight into specific mechanisms used by a TF to recognize DNA targets (2–5). One example for highly specific H-bonds occurs with arginine, which recognizes guanine through a bidentate interaction—forming two adjacent H-bonds—often contributing significantly to TF–DNA binding specificity (4, 6). In addition, protein residues can employ hydrophobic interactions to differentiate nucleotides, for example thymine versus cytosine (7–9). In some cases, structural deviations from a B-form double helix increase accessibility of DNA physicochemical signatures, enabling a TF to establish an optimized set of H-bonds or hydrophobic interactions that can determine DNA-binding specificity (10–12). Nevertheless, there is a paucity of experimentally determined TF structures in complex with DNA target sites. Available structural information for TF–DNA complexes is typically limited to complexes where a protein or its DNA-binding domain binds to a single DNA sequence.

In the last decade, several high-throughput binding technologies have been developed to enable sampling of TF–DNA binding. These technologies quantitatively measure the binding affinities of one TF against thousands or even millions of different DNA sequences in vitro (13–21). These methods provide an alternative path to infer TF–DNA binding without requiring time-consuming structural biology experiments. Concurrently, several DNA motif discovery methods have been established for modeling TF–DNA binding preferences using these high-throughput experimental data. Methods based on position weight matrices (PWMs) assume that each nucleotide at a corresponding position

Significance

The Watson–Crick sequence model enables a simplified representation of DNA, wherein four letters, A, C, G, and T, describe the chemical identities and orientations of all possible nucleotide pairs. In this coarse-grained model, each letter describes an assembly of over 60 atoms. However, the atomic composition of a nucleotide pair can be altered by chemical modifications, different base-pairing geometries, or mismatches. As only a few atoms contribute to binding specificity, we propose that compared to a sequence model, a chemistry-based model that directly encodes protein–DNA contacts may more robustly capture the chemical variations of DNA. We introduce models that directly and precisely represent physicochemical readout, which is, importantly, not restricted to standard Watson–Crick base pairs.

[1]Present address: Department of Biochemistry and Molecular Genetics, University of Colorado Denver School of Medicine, Aurora, CO 80045.

[2]To whom correspondence may be addressed. Email: rohs@usc.edu.

independently contributes to the overall binding affinity (22, 23). To describe interdependent contributions between nucleotides, more complex modeling methods have been developed (24–26). Nevertheless, none of these methods directly infers chemical interactions of amino acids in DNA grooves that are essential for binding specificity.

Physicochemical signatures of conventional DNA bases (Fig. 1C) can be altered by modifying a chemical group. These signatures constitute an important layer of reprogrammable information in DNA (Fig. 1D). For example, the chemical signatures of 5mCpG dinucleotides where cytosine is methylated at the 5-carbon position can be recognized by the methyl-CpG-binding domain protein that recruits histone deacetylases and promotes local chromatin condensation to regulate transcription (27). Non-Watson–Crick bp, such as Hoogsteen (6, 28) (Fig. 1E), synthetic (29) (Fig. 1F), and mismatched (30) (Fig. 1G) bp, represent unique physicochemical signatures in DNA. These unique bp introduce new layers of complexity and possibly influence TF–DNA recognition. For example, mismatched DNA can be recognized by a specific class of TFs that acts as a repair barrier to increase the mutation rate and thereby regulate cellular replication and repair processes (30–32). Mismatched bp have recently also been reported to play a role in the CRISPR-Cas9 gene editing system (33). Investigating how these non-standard physicochemical signatures affect binding specificity is an important step toward understanding the binding mechanisms. However, existing

methods are difficult to apply to DNA modifications or non-Watson–Crick bp, due to a potential overfitting problem as a result of one-hot encoding with expanded alphabets representing such non-standard bp.

Here, we introduce DeepRec (Deep Recognition for TF–DNA binding), a deep-learning-based method that integrates two convolutional neural network (CNN) modules for extracting the pattern of physicochemical signatures in the major and minor grooves of DNA. Each CNN module extracts nonlinear spatial context among physicochemical signatures of bp to mine potential insights beyond DNA sequence. We use a grid hyperparameter search to find a combination of hyperparameters, which yields an optimal model to minimize a predefined loss function on a given dataset. To reduce the error introduced by an individual predictive model, we performed ensemble training with multiple random seeds and average the contribution of each physicochemical signature. DeepRec integrates a forward perturbation-based interpretative approach that highlights the important physicochemical signatures for deciphering binding mechanisms. This method aims to reveal important physicochemical patterns recognized by TFs and further explain biological insights that cannot be elucidated by sequence-based models. Such a chemistry-based approach is necessary, given the increasing evidence for the biological importance of various chemical modifications of DNA in gene regulation, cellular function, and disease.
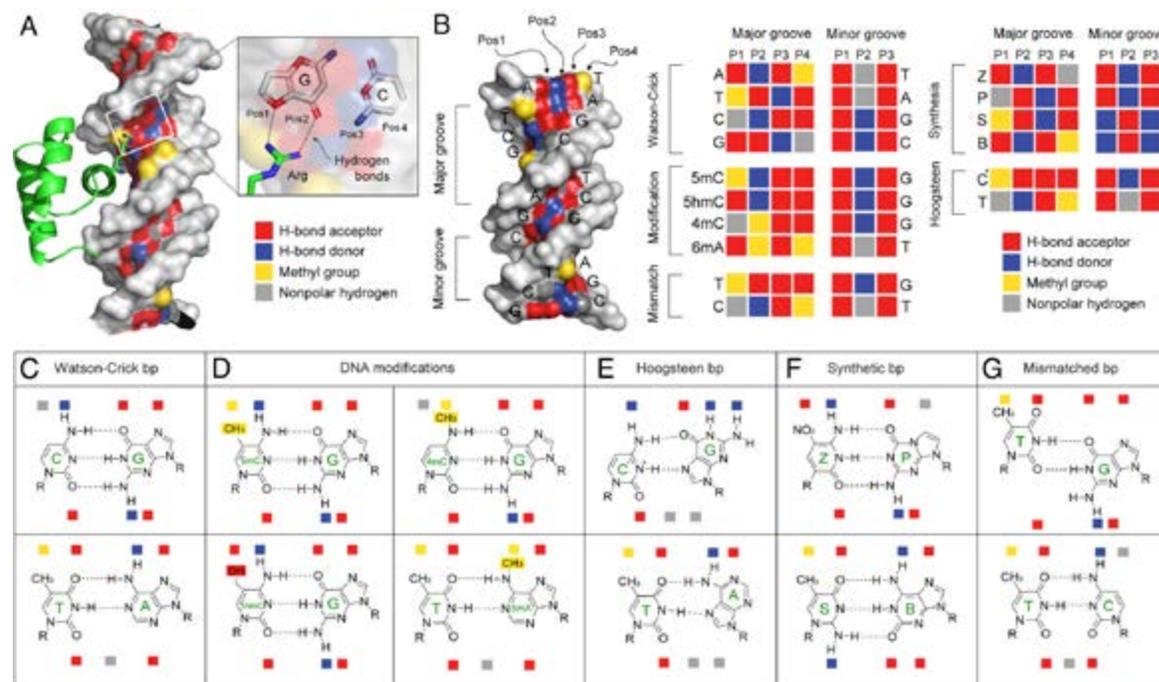

Fig. 1. Physicochemical signatures of base pairs (bp) characterized by a set of chemical groups at bp edges in the major or minor groove. (A) Schematic showing that the DNA binding domain of a TF recognizes DNA physicochemical signatures, including an H-bond acceptor (red), H-bond donor (blue), methyl group (yellow), and nonpolar hydrogen (gray). (B) Schematic view demonstrating the standard and expanded physicochemical signatures in DNA major and minor grooves. Labels 'Pos1-4' (major groove) and 'Pos1-3' (minor groove) indicate positions of physicochemical groups at various bp. In addition to (C) standard Watson–Crick bp signatures, additional unique signatures are introduced by (D) DNA modifications, (E) Hoogsteen bp, (F) synthetic nucleotides, and (G) mismatched bp. 5-methylcytosine (5mC)/G and 5-hydroxymethylcytosine (5hmC)/G modifications carry unique signatures in the major groove by an additional methyl or hydroxymethyl group on the 5-position of cytosine, respectively. Methylations on the 4-position of the cytosine pyrimidine ring (N4-methylcytosine, 4mC) and on the 6-position of the adenine purine ring (N6-methyladenine, 6mA) create alternate unique signatures in the major grove. Hoogsteen bp flip their purines and exhibit different physicochemical patterns in the major or minor groove. Hachimoji DNA is composed of four synthetic bases (P, Z, S, B). The Z/P bp has a unique pattern in the major groove, whereas the S/B bp has a unique pattern in the minor groove. Mismatched bp can change the number of H-bonds in the base-pairing geometry. For example, the T/G bp increases by one H-bond, and the T/C bp loses one H-bond acceptor.

Results and Discussion

Description of the DeepRec Framework. The framework of DeepRec consists of two parallel CNN modules to extract binding patterns from physicochemical signatures in the DNA major and minor groove, respectively (Fig. 2*A*). A joint layer combines patterns extracted from the two CNN modules, and a hidden layer further discovers higher-level binding patterns (Fig. 2*A*). The model is capable of characterizing the relative affinities of all binding sites by building a predictive model based on high-throughput experimental data.

We introduce a new encoding method that has a contact-based representation in the context of TF–DNA recognition in comparison to the widely adopted one-hot encoding method for DNA sequences (A, C, T, and G). Our method encodes physicochemical signatures in the major and minor grooves with respect to their defined positions into two-dimensional features or images that contain geometric relationships among physicochemical signatures (Fig. 1*B*) (1). The CNN modules keep the original structure of the input and extract the spatial context of the physicochemical signatures to identify binding patterns. This information might be neglected by sequence-based models that consider DNA as a one-dimensional string of letters, likely assuming that the features of DNA bp are independent of each other. The hidden layer models the higher-order interactions between patterns implicitly. In this way, our method is able to mine spatial and nonlinear information from the physicochemical space.

Compared to the one-hot sequence-encoding scheme, our method provides a fundamental description of the actual molecular interactions between TFs and DNA, and it retains the dependency between bp based on the physicochemical

signatures at their major and minor groove edges. For example, the only difference between cytosine and 5-methylcytosine (5mC) is a methyl group at one major-groove position (Fig. 1*B*). One-hot sequence encoding assumes that all nucleotides are independent, which might lead to loss of information about nucleotide interdependence. Specifically, the dependency of unmethylated, fully methylated, and hemimethylated CpG bp steps varies in different degrees. Such variable dependency cannot be represented by independent letters in a one-hot sequence-encoding scheme.

Another advantage of our new encoding scheme is that the encoding of physicochemical signatures does not need to be expanded when introducing diverse nucleotide types, such as those from chemical DNA modifications, or from Hoogsteen (6, 28), synthetic (29), or mismatched bp (30). By contrast, the one-hot sequence-encoding scheme must expand the feature dimension, and a massive yet sparse encoding matrix is more likely to lead to an overfitting issue when modeling the data.

We use deep learning techniques to infer model parameters. Our training pipeline alleviates the need for manual parameter adjustment by automatically tuning several calibration parameters through threefold cross-validation (Fig. 2*B*). With the tuned parameters, we perform ensemble training with multiple random seeds and filter out low-performance models (Fig. 2*C*). DeepRec utilizes a forward perturbation-based approach to calculate the binding difference between the presence and absence of a chemical group at a given position, regardless of nucleotide sequence (Fig. 2*D*). To visualize the binding preferences of an individual TF, DeepRec introduces a new visualization of physicochemical signatures using physicochemical energy logos (Fig. 2*E*). The package is available on GitHub (https://github.com/TsuPeiChiu/DeepRec).
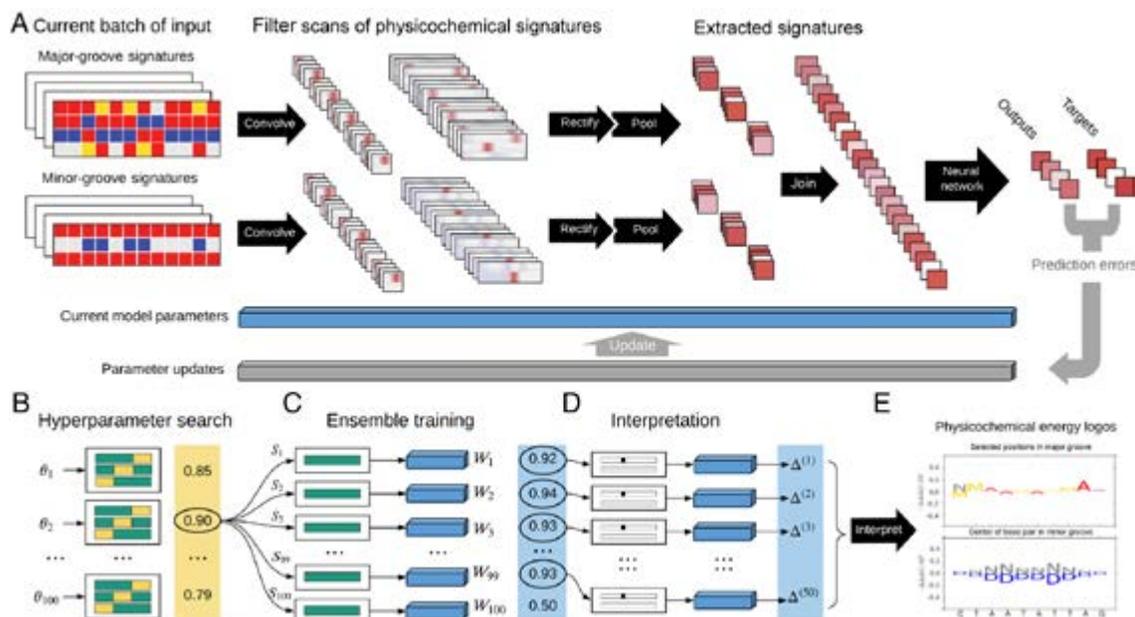

Fig. 2. Workflow of the DeepRec prediction framework. (*A*) The stages of convolution, rectifying, pooling, and neural networks predict binding affinities based on major and minor groove positions (Pos) for each input sample. During the training phase, back-propagation stages simultaneously update all trainable parameters to improve prediction accuracy. The entire process starts from (*B*) a hyperparameter search that randomly chooses 100 parameter combinations in hyperparameter space θ and performs threefold cross-validation on the training set. The parameter combination with the highest average r^2 is used for (*C*) ensemble training with 100 random seeds (S_1 to S_{100}) and results in 100 models (W_1 to W_{100}). Models of the top 50 performers in r^2 are tested on held-out validation sets and used for (*D*) interpretation. The delta value (Δ) is the gap between relative binding affinities when a specific physicochemical signature is or is not present. (*E*) The average of delta values (Δ_1 to Δ_{50}) over the 50 models is used for calculation of the physicochemical energy logos.

In this study, we demonstrated and validated our physicochemical encoding scheme of DNA for the TFs MAX, MEF2B, p53, ATF4, and C/EBPβ. We selected these TFs based on the primary consideration that they have representative binding mechanisms, and they each have available high-quality high-throughput experimental data and corresponding co-crystal structures. In terms of binding mechanisms, MAX binds to the DNA major groove with several types of H-bonding (34). MEF2B is known for its essential minor groove binding (35). p53 recognizes its DNA binding site as a protein tetramer and through an interplay of base and shape readout (6). Finally, ATF4 binds to a CpG dinucleotide through the mechanism of thymine mimicry (36). All these systems have SELEX-seq experimental data that characterize a full range of TF-binding affinities for model training and corresponding co-crystal structures in the Protein Data Bank (PDB) for validation (37).

DeepRec Predicts DNA Contacts in TF-DNA Binding. Experimentally determined structures of TF–DNA complexes provide critical insights into binding mechanisms. By observing the number and geometry of H-bonds as well as hydrophobic contacts between protein residues and bp, one can understand how proteins use their unique readout mechanisms to achieve DNA-binding specificity. However, co-crystal structures are available for relatively few TFs and are typically limited to complexes in which a protein or its DNA-binding domain binds to a single DNA sequence. Moreover, crystal-packing contacts near the binding site, variations of side-chain rotamers, and the presence of low-electron-density polar hydrogens can limit attempts to identify a binding contact. DeepRec leverages large data generated from high-throughput binding

assays and enables the prediction of physicochemical readout from sequencing data. In this way, DeepRec has the potential to confirm and reveal unknown binding mechanisms without the requirement of solving a structure.

We first targeted the widely studied human helix–loop–helix (bHLH) protein MAX, which preferentially recognizes a subset of enhancer-box (E-box) sequences (38). We trained DeepRec on MAX SELEX-seq data (39), interpreted the model using DNA physicochemical energy logos (Fig. 3A), and compared predicted logos with sequence logos (Fig. 3B) and TF contacts within a MAX–DNA co-crystal structure (PDB ID 5EYO) (34) (Fig. 3 C–F). We observed two prominent H-bond acceptors (denoted as 'A's) at physicochemical signature position 3 (Pos3) and Pos4 in the major groove of the C_{-3}/G_{-3} bp (or at Pos1 and Pos2 of the G_3/C_3 bp) in the E-box of the physicochemical energy logos (Fig. 3A). The logos showed a positive average change in binding free energy $-\Delta\Delta G$ (see definitions in *SI Appendix*). Consistent with the co-crystal structure of the MAX–DNA complex, His28 forms one H-bond with either O_6 or N_7 of the 3′ guanine (G_{-3}), or forms a bifurcated H-bond (with the two H-bonds sharing a donor) (1) (Fig. 3C). We observed no clear signal in energy logos at Pos2 in the major groove of the C_{-3}/G_{-3} bp (or Pos3 of G_3/C_3) (Fig. 3A). In the MAX–DNA co-crystal structure (Fig. 3C), the H-bond acceptor of Glu32 might be occupied by donors of Arg35, preventing bonds with the donor of the 5′ cytosine (C_{-3} or C_3).

Another notable H-bond acceptor was found in the physico-chemical energy logos at Pos3 in the major groove of the A_{-2}/T_{-2} bp (or Pos2 of the T_2/A_2 bp) (Fig. 3A). This observation is consistent with the co-crystal structure, in which Arg36 donates one or two H-bonds to O_4 of the 3′ thymine (T_{-2} or T_2) (Fig. 3D). One
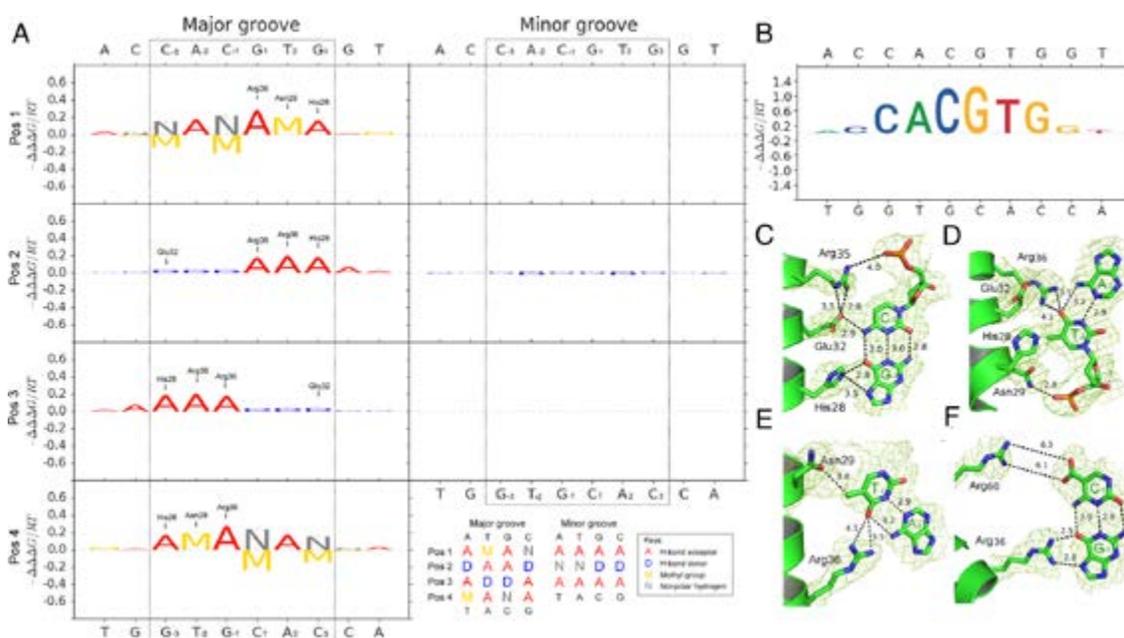

Fig. 3. DNA physicochemical energy logos and co-crystal structure of the MAX–5caC complex (PDB ID 5EYO). (A) Physicochemical energy logos were generated by DeepRec based on MAX SELEX-seq data. The physicochemical signature position (Pos) and the nucleotide position of the bp are indicated by the name of the MAX protein residue when a contact between the protein and DNA was found, using the DNAproDB method (26). Pos1-4 refer to physicochemical signature positions 1-4, respectively, of the bp. The dashed box highlights the MAX binding site, and the sequence on top of these panels indicates the input sequence. In the energy logos, 'M' represents the thymine methyl group, since the model was trained on unmethylated data. (B) Sequence logos obtained from DeepRec sequence model. (C) Interactions between MAX residues and the C_{-3}/G_{-3} bp in the major groove. Numbers indicate distances between atoms in Å. (D and E) Interactions between MAX residues and the A_{-2}/T_{-2} bp in the major groove. (F) Interactions between MAX residues and the C_{-1}/G_{-1} bp in the major groove.

notable methyl-group was found at Pos4 in the major groove of the A_{-2}/T_{-2} bp (or Pos1 of the T_2/A_2 bp). In the co-crystal structure, Asn29 formed a van der Waals (vdW) interaction with the methyl-group of the 3′ thymine (T_{-2} or T_2) (Fig. 3E). Two additional prominent H-bond acceptors were found at Pos3 and Pos4 in the major groove of the C_{-1}/G_{-1} bp (or Pos1 and Pos2 of the G_1/C_1 bp) (Fig. 3A). In the co-crystal structure (Fig. 3F), Arg36 interacts with the central 5′-CpG-3′ dinucleotide by donating two hydrogen atoms to the O_6 and N_7 atoms of the 3′ guanine (G_{-1} or G_1), forming a bidentate H-bond (two H-bonds with different donor and acceptor atoms). This geometry conveys a high degree of specificity (40). Similar results were obtained when we cross-validated our method with data from the microfluidics-based SMiLE-seq platform (21) (*SI Appendix*, Fig. S1). To further prove that our method is generalizable across different experimental platforms, we predicted relative binding affinities measured by SMiLE-seq with the model trained by the SELEX-seq data. The results showed a good correlation (*SI Appendix*, Fig. S2). As a control, we shuffled the relative binding affinities of the training data, and no signals were detected from the predictive model (*SI Appendix*, Fig. S3). Together, these results showed that DeepRec is capable of predicting binding mechanisms with respect to the co-crystal structure.

DeepRec Predicts TF Binding Preference in the Context of DNA Grooves. Physicochemical signatures in the DNA major groove are more diverse and unique than those in the minor groove, thereby conveying a higher degree of TF-binding specificity (1). For example, MAX only recognizes signatures in the major groove to achieve its binding specificity, as evidenced by the physicochemical energy logos (Fig. 3A) and the MAX–DNA co-crystal structure (PDB ID 5EYO). Nevertheless, TF-binding specificity can be

achieved through a complex recognition process involving both major and minor groove readout (1). Therefore, we asked whether DeepRec is able to predict binding mechanisms that involve physicochemical signatures in the major and minor grooves.

We next studied myocyte enhancer factor 2B (MEF2B), a member of the MEF2 family that plays vital roles in the development and functioning of neuronal and muscle cells. Minor groove contacts have been shown to be important for TF binding for this family (41). We trained DeepRec on MEF2B SELEX-seq data (41), interpreted the model using DNA physicochemical energy logos (Fig. 4A), and compared predicted logos with the sequence logos (Fig. 4B) and the TF–DNA contacts of a MEF2B–DNA co-crystal structure (PDB ID 1N6J) (42) (Fig. 4 C–F). One prominent H-bond acceptor at Pos4 in the major groove of the T_{-4}/A_{-4} bp (or Pos1 in the major groove of the A_4/T_4 bp) was observed in the physicochemical energy logos (Fig. 4A). Consistent with the MEF2B–DNA co-crystal structure, Lys23 donates one H-bond to N_7 of the 3′ adenine (A_{-4} or A_4) (Fig. 4D). Intriguingly, the major groove signals mainly occurred at the 3-bp half sites (5′-CTAW$_4$TAG-3′). In contrast, notable minor groove signals were observed within the central W_4 region of the binding site (Fig. 4 A, *Right*).

Compared to the physicochemical energy logos of MAX (Fig. 3 A, *Right*), MEF2B logos showed stronger signals in the minor groove, with positive nonpolar hydrogens ('N's) and negative H-bond donors ('D's) (Fig. 4 A, *Right*). In the MEF2B–DNA co-crystal structure (Fig. 4E), the Gly2-Arg3 conformation inserts into the minor groove: Arg3 makes electrostatic interactions with the phosphodiester backbone, whereas Gly2 makes vdW or hydrophobic interactions with nonpolar atoms. The energy logos demonstrated long-range interactions in the minor
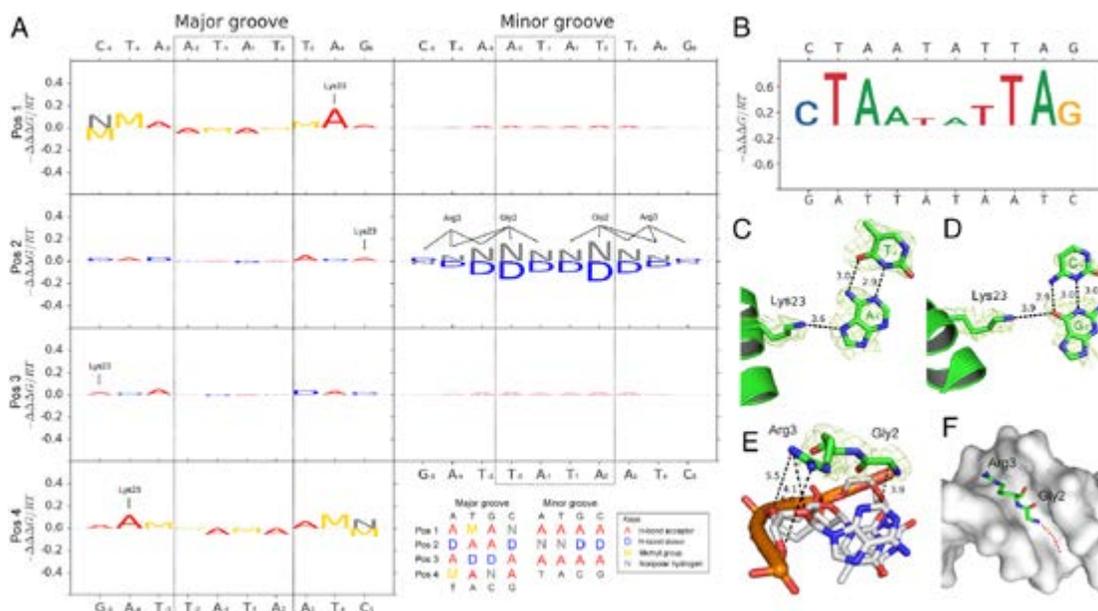

Fig. 4. DNA physicochemical energy logos and co-crystal structure of MEF2B–DNA complex (PDB ID 1N6J). (*A*) Physicochemical energy logos were generated by DeepRec based on MEF2B SELEX-seq data. The position with respect to the physicochemical signature position (Pos) and the nucleotide position of the bp are indicated by the name of the MEF2B protein residue when a contact between the protein and DNA was found, using the DNAproDB method (26). Pos1-4 refer to physicochemical signature positions 1-4, respectively, of the bp. The dashed box highlights the central core-binding site. (*B*) Sequence logos obtained from DeepRec sequence model. (*B*) Interactions between MEF2B residues and the C_{-5}/G_{-5} bp in the major groove. Numbers indicate distances between atoms in Å. (*C*) Interactions between MEF2B residues and T_{-4}/A_{-4} bp in the major groove. (*D*) Interactions between MEF2B residues and the T_{-5}/A_{-5} bp in the major groove. (*E*) Interactions between MEF2B residues and atoms in the minor groove. (*F*) "Met-Gly-Arg" conformation in the minor groove.

groove (Fig. 4 *A*, *Right*), suggesting the potential presence of another nonpolar residue such as Met, which can form highly specific interactions for minor-groove recognition (Fig. 4*F*).

DeepRec Predicts the Geometry of H-Bonding and Hydrophobic Interactions. DNA-binding specificity is achieved through H-bond contacts and hydrophobic interactions. While the number of contacts formed between protein residues and DNA bases provides binding specificity, the uniqueness of the geometry of H-bonds and the hydrophobic interaction conveys a higher degree of specificity for protein recognition (1). For example, the H-bond geometry of bidentate and bifurcated H-bonds provides a higher degree of binding specificity than single H-bonds (43). In a complex example, bidentate H-bonds combined with a hydrophobic interaction form a triad geometry that conveys an even higher degree in binding specificity. For instance, a methyl-Arg-G triad forms during recognition of TpG and methyl-CpG dinucleotides in double-stranded DNA (44).

The human tumor suppressor p53 binds as a tetramer to two dimeric sites (5′- RRRCWWGYYY -3′) separated by a spacer of 0-13 bp (45). Each p53 monomer uses a methyl-Arg-G triad to recognize the TpG dinucleotide (44). We trained DeepRec on p53 SELEX-seq data (39), interpreted the model using DNA physicochemical energy logos (Fig. 5*A*), and compared predicted logos with the sequence logos and TF–DNA contacts of the p53–DNA co-crystal structure (PDB ID 3Q06) (1) (Fig. 5 *B–D*). We observed two prominent H-bond acceptors at Pos1 and Pos2 in the major groove of the G_2/C_2 bp in the physicochemical energy

logos (Fig. 5*A*). This observation is consistent with the p53–DNA co-crystal structure, in which Arg280 donates two H-donors to N_6 and O_7 of the 3′ guanine, forming bidentate H-bonds (Fig. 5*D*).

On the other hand, a prominent methyl group was found at Pos1 in the major groove of the T_1/A_1 bp (Fig. 5*A*), in agreement with the co-crystal structure wherein the methyl group of thymine stabilizes the Arg280 positioning through a vdW contact (Fig. 5*D*). The bidentate H-bond and the hydrophobic interaction formed a so-called "methyl-Arg-G triad," providing unique specificity. Intriguingly, three consecutive 'A's on the side are found at Pos1 in the major groove of G_{-5}, G_{-4}, and A_{-3}, near H-donor Arg280 and Lys120, suggesting why p53 prefers the RRR triplet at the binding-site edges (Fig. 5*C*). The results showed that DeepRec is capable of detecting the binding geometry for p53–DNA recognition.

DeepRec Predicts Impact of DNA Modifications on Protein–DNA Binding. DNA modifications play key roles in gene regulation (36, 46), but their effects on TF-binding are not completely understood. Several studies have adopted structural biology and low-throughput binding methods (e.g., X-ray crystallography and electrophoretic mobility shift assays) on TFs in complexes with modified DNA to explain the effects of chemical modifications at atomic resolution (34, 47). Recently, high-throughput methods, such as EpiSELEX-seq (36), methyl-HT-SELEX (46), and methyl-Spec-seq (48), have been developed to determine the effects of the CpG modification on TF-binding using a large pool of
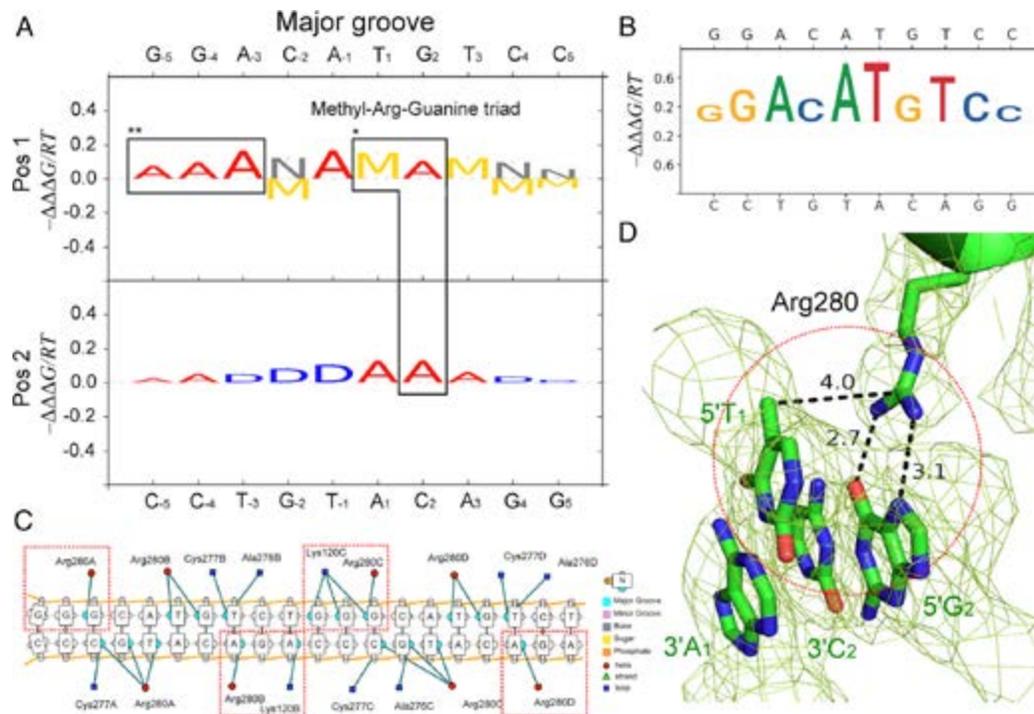

Fig. 5. Selected DNA physicochemical energy logos and co-crystal structure of p53–DNA complex structure (PDB ID 3Q06). (*A*) Preferred physicochemical signature position (Pos) indicates preference for the thymine methyl group and two guanine acceptors in the major groove of the p53 response element. Pos1-2 refer to physicochemical signature positions 1-2, respectively, of the bp. Positions marked with * represent Arg280 contacts corresponding to (*D*) the p53–DNA co-crystal structure. (*B*) Sequence logos obtained from DeepRec sequence model. (*C*) TF–DNA contacts in red dashed-boxes show preference of purine bases due to H-bonds between the acceptors and the donors provided from arginine and lysine. Contact map is obtained from DNAproDB. Contact map is consistent with the geometrical pattern shown in positions marked with ** in panel (*A*).

chemically modified oligonucleotides. These high-throughput methods are powerful tools but lack insights into structural readout mechanisms.

Here, we asked whether we could predict the effects of methylation on TF binding using high-throughput binding data and explain possible binding mechanisms. We trained DeepRec on EpiSELEX-seq data (36) of the human bZIP proteins ATF4 and C/EBPβ using a similar process as was used in previous systems. In this case, we combined methylated (Lib-M) and unmethylated (Lib-U) DNA fragments generated from EpiSELEX-seq (36) as input. We modeled the systems without expanding the sequence alphabets by instead describing physicochemical signatures (Fig. 1 *B* and *D*). Predictions for ATF4 showed a decrease in binding affinity when a CpG bp step is present at the center of the binding site (at position –1/+1), and an increase in binding affinity when

the sequences contained a CpG dinucleotide at the flank of the motif (at position –4/–3 or +3/+4) (Fig. 6*A* and *SI Appendix,* Fig. S3*A*). In contrast, predictions for C/EBPβ showed a weak or no preference for 5mCpG at the center of the binding site (Fig. 6*B* and *SI Appendix,* Fig. S3*B*) (49). The methylation effect was precisely identified in low-affinity binding sequences for ATF4 (Fig. 6*A* and *SI Appendix,* Fig. S3*E*). These results are consistent with results from previous studies (*SI Appendix,* Fig. S3 *C–F*) (36, 46, 49, 50).

Next, we trained DeepRec on SELEX-seq data (39) for ATF4 and C/EBPβ. We asked whether we could learn the methylation effect from unmethylated DNA. Predictions for ATF4 showed a positive effect on binding affinity when the sequences contained a CpG bp step in the flank, demonstrating the 'thymine mimicry' that could possibly be learned from the methyl group of thymine
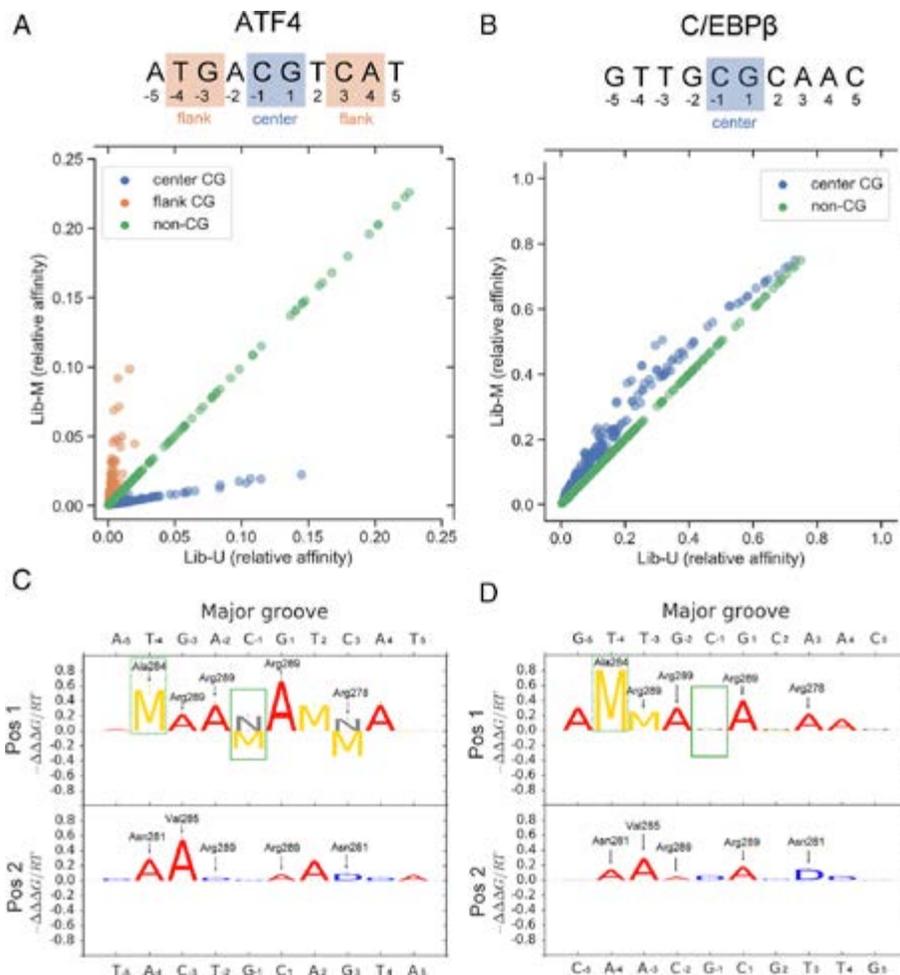

Fig. 6. Predicting and measuring methylation sensitivity for ATF4 and C/EBPβ using DeepRec on EpiSELEX data and validating the contacts with C/EBPβ complex structure (PDB ID 6MG1 and 2E42). Symmetric consensus motifs and comparison of relative affinities of 10-mer sequences between a methylated sequence library (Lib-M) versus an unmethylated sequence library (Lib-U) predicted by DeepRec are shown for (*A*) ATF4 and (*B*) C/EBPβ. For ATF4 (*A*), non-CpG sequences (green) show the same binding affinities in both libraries because no methylation is involved. Flanking CpG-containing sequences are preferred in Lib-M, whereas central CpG-containing sequences are preferred in Lib-U. For C/EBPβ (*B*), central CpG-containing sequences are slightly preferred in Lib-M. DNA physicochemical energy logos are shown for (*C*) ATF4 and (*D*) C/EBPβ. Corresponding contacts from co-crystal structures (PDB ID 6MG1) are indicated by residue name and number. Due to the unavailability of a ATF4 co-crystal structure, the C/EBPβ co-crystal structure was compared between ATF4 logos and contacts marked by a dashed line. Green dashed-boxes highlight hydrophobic interactions between the methyl group of the thymine and the hydrophobic side chains that are shared by ATF4 and C/EBPβ. Green solid boxes indicate a negative effect of methylation on central CpG for ATF4 but a slight positive effect for C/EBPβ.

in the unmethylated data (*SI Appendix,* Fig. S3 *G* and *I*). However, the model trained for C/EBPβ did not capture the positive contribution toward binding from the thymine in the central TpG dinucleotide (*SI Appendix,* Fig. S3 *H* and *J*), which implied that the thymine mimicry event might not be selected in this case. This comparison allowed us to study the possible mechanisms of the methylation effect on binding.

The resulting physicochemical logos supported the previously identified molecular basis for methylation effects on binding (36). The predictions showed a positive 'M' at Pos1 in the major groove of T_{-4}/A_{-4} (Pos4 in the major groove of A_4/T_4) for ATF4 (green dashed boxes in Fig. 6 *C* and *D*), demonstrating the positive effect of methylation on binding affinity. Based on TF–DNA contacts in a C/EBPβ co-crystal structure (PDB ID 6MG2) (51), Ala284 interacts with the methyl group of a thymine at position –4 of the binding site, making a vdW contact Fig. 6*E*. A negative 'M' is found at Pos1 in the major groove of C_{-1}/G_{-1} (Pos4 in the major groove of G_1/C_1) for ATF4 (Fig. 6*C*), showing a negative effect on binding affinity. However, there is no obvious signal of 'M' detected at Pos1 in the major groove of C_{-1}/G_{-1} (or at Pos4 in the major groove of G_1/C_1) for C/EBPβ (Fig. 6*D*). In the co-crystal structure Fig. 6*H*, the 'M' of an 5mC might interact with the carbon Arg289 to form a vdW interaction, which may explain how C/EBPβ interacts with methylated DNA Fig. 6*H*. However, why ATF4 has less preference for methylated DNA is still unknown.

Conclusions

Investigating physicochemical readout signatures in DNA that are important for protein binding is an approach to uncover TF–DNA readout mechanisms. However, current experimental methods are limited in their ability to provide sufficient numbers of structures to explain the entirety of TF–DNA binding mechanisms. Until now, existing methods have been incapable of mining important physicochemical signatures by leveraging large data. Here, we describe DeepRec, a deep learning framework capable of mining the importance of physicochemical readout signatures for TF–DNA binding specificity in the context of binding free energy. DeepRec accurately predicts possible binding contacts for several TFs across TF families at physicochemical resolution, and indicates corresponding forces (e.g., H-bonds, hydrophobic interactions, etc.) that current sequence-based modeling methods cannot identify. Furthermore, DeepRec detects physicochemical signatures and binding geometries that can contribute to highly specific TF–DNA binding, such as bifurcated and bidentate H-bonds, methyl-Arg-G triads, and long-range-DNA minor groove interaction patterns.

Strikingly, DeepRec is capable of revealing effects of chemical DNA modifications on learning on a dataset that combines methylated and unmethylated data. Compared with results trained on unmethylated data alone, one can imply the possible binding mechanism of a methylation event. Because DeepRec can mine lesser coarse-grained information than DNA sequence, we envision that this method can be easily expanded to studies beyond the concept of DNA sequence towards physicochemical modeling. For example, DeepRec might be employed to investigate effects of Hoogsteen bp (52) observed in p53 (6), as well as effects of synthetic (29), mismatched bp (30), or other modified bp, such as 4mC and 6mA (53), in the context of TF–DNA binding.

Future work can improve the DeepRec approach. First, we currently limit the number of physicochemical signature positions in major and minor grooves to four and three, respectively; however, in some cases this definition might be vague. For example,

the number of physicochemical signature positions in the minor groove could be two or three. In addition, we encode physicochemical signatures as discrete data and consider O and N as the same type of physicochemical signatures with equal physicochemical property strength. However, considering the strength of the acceptor from a hydroxymethyl group, O, and N would be different. To address this issue, we should introduce a function to describe the distribution of strength for different chemical groups rather than use a discrete one-hot encoding representation. Furthermore, DeepRec requires large data containing enough variance to be able to train thousands of parameters. Some data, such as HT-SELEX, may not be ideal for our method. Compared to our end-to-end method, a pipeline training from a sequence model or a conversion from PWM to obtain physicochemical logos might neglect the important spatial context of physicochemical signatures. The content-switch process would introduce more arbitrary parameters. DeepRec trains models with shape features implicitly; therefore, the impact of three-dimensional structure could not be highlighted by our method. Finally, DeepRec might not be able to predict all contacts occurring in the co-crystal structure, or vice versa. To improve the prediction, we could co-train our model with existing crystal structures to fine-tune the predictions.

Materials and Methods

Deep Learning Framework DeepRec. Deep neural networks are a type of artificial neural networks comprised of multiple layers between input and output layers. Each layer consists of a number of neurons, which receive input from a set of previous layer neurons. This sequential layer-by-layer structure executes a sequence of functional transformations to model complex nonlinear relationships between predictive features and response variables.

We developed a multimodule deep-learning framework capable of mining important patterns in multimodal systems. The architecture includes convolutional and down-sampling layers for each module to extract features from input data, a joint layer combining features retrieved from different sources, and a hidden layer that further integrates features to discover higher-level patterns. Based on this framework, we developed DeepRec (<u>Deep Rec</u>ognition of TF–DNA binding), which integrates physicochemical features of DNA in the major and minor grooves, followed by a perturbation-based forward-propagation approach to interpret the resulting model (Fig. 2). This method aims to discover important physicochemical readout signatures recognized by TFs and to explain biological insights that cannot be revealed by sequence-based models.

DNA Physicochemical Signatures and Feature Encoding. Physicochemical signatures at the edges of bp in the DNA major or minor grooves underlie the ability of TFs to recognize bp through H-bonds or hydrophobic contacts, as shown in Fig. 1. For a given DNA sequence, we encode the corresponding physicochemical signatures using a binary representation for H-bond acceptor, H-bond donor, thymine methyl group, and nonpolar hydrogen. A detailed description about the encoding method is provided in *SI Appendix*. The encoding method can be extended to non-Watson–Crick bp, including Hoogsteen, synthetic, and mismatched bp, without increasing the feature dimension. In contrast, the sequence-based model introduces entirely new letters of the sequence alphabet when using the one-hot encoding method, which would increase the dimension of input features by making them sparse, which might result in an overfitting issue. Using a different letter also implies independence, for instance of a methylated cytosine from cytosine despite the largely overlapping chemical characteristics of C/G and 5mC/G bp.

Hyperparameter Search. The hyperparameter search begins by sampling 100 sets of random calibration parameters. *SI Appendix,* Table S1 lists the sampling used for each parameter in a set. The calibration phase evaluates the quality of each parameter set by threefold cross-validation on the training set. Each model is trained on a different two-thirds of the data, and its performance is evaluated on the held-out one-third. Calibration parameters are scored by averaging the three r^2 values of the validation dataset (Fig. 2*B*).

Ensemble Training. Once the best calibration parameters have been identified, we trained 100 new models using 80% of the training data (a single fold) with different random seeds. Resulting models are stored and used to reduce the variance of predictions and generalization errors that generally happen in neural networks. The best 0.5 quantile of models is selected and is returned by the entire pipeline for interpretation or further analysis.

Model Interpretation. DeepRec utilizes a perturbation-based forward propagation approach that nullifies a physicochemical signature at each defined position of a bp, one at a time, and then quantifies its impact on binding free energy. The binding free-energy difference is calculated between the presence and absence of the signature at each corresponding physicochemical signature position. To visualize the detailed binding preferences of an individual TF, DeepRec introduces a new visualization, coined the 'DNA physicochemical energy logo' (e.g., Fig. 3*A*). In these logos, letters are used to represent DNA physicochemical features ('A' for H-bond acceptor, 'D' for H-bond donor, 'M' for methyl group, and 'N' for nonpolar hydrogen). The logos describe the binding preference in each DNA groove (major and minor) and physicochemical group position 1-4 (Pos1-4) at single-nucleotide resolution. The height of each letter indicates the average change in binding free energy ($-\Delta\Delta G$) resulting from the comparison of the reference probe to its mutants with a nullified physicochemical signature in ensemble models. The vertical bar on each letter specifies the standard error of the mean, which measures how far the sample mean is likely to be separated from the true population mean.

This method is expanded to handle DNA modifications by considering the addition or replacement of a specific physicochemical signature. For example, the 'N' signature at Pos1 of the C/T bp can be swapped with an 'M' signature in the major groove to represent methylation at position 5 of the cytosine, 5mC. Next, the change in binding free energy can be measured by comparing the reference probe with nullification of the physicochemical signature.

Data, Materials, and Software Availability. Previously published data were used for this work (PRJEB25690; SRP073361; GSE116401; GSE98652).

Author affiliations: ªDepartment of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA 90089; ᵇDepartment of Chemistry, University of Southern California, Los Angeles, CA 90089; ᶜDepartment of Physics and Astronomy, University of Southern California, Los Angeles, CA 90089; and ᵈDepartment of Computer Science, University of Southern California, Los Angeles, CA 90089

1. R. Rohs *et al.*, Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* **79**, 233–269 (2010).
2. J. M. Sagendorf, N. Markarian, H. M. Berman, R. Rohs, DNAproDB: An expanded database and web-based tool for structural analysis of DNA-protein complexes. *Nucleic Acids Res.* **48**, D277–D287 (2020).
3. R. Joshi *et al.*, Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* **131**, 530–543 (2007).
4. C. W. Garvie, C. Wolberger, Recognition of specific DNA sequences. *Mol. Cell* **8**, 937–946 (2001).
5. N. M. Luscombe, S. E. Austin, H. M. Berman, J. M. Thornton, An overview of the structures of protein-DNA complexes. *Genome Biol.* **1**, REVIEWS001 (2000).
6. M. Kitayner *et al.*, Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. *Nat. Struct. Mol. Biol.* **17**, 423–429 (2010).
7. S. C. Harrison, A. K. Aggarwal, DNA recognition by proteins with the helix-turn-helix motif. *Annu. Rev. Biochem.* **59**, 933–969 (1990).
8. A. K. Aggarwal, D. W. Rodgers, M. Drottar, M. Ptashne, S. C. Harrison, Recognition of a DNA operator by the repressor of phage 434: A view at high resolution. *Science* **242**, 899–907 (1988).
9. C. Wolberger, Y. C. Dong, M. Ptashne, S. C. Harrison, Structure of a phage 434 Cro/DNA complex. *Nature* **335**, 789–795 (1988).
10. R. S. Hegde, S. R. Grossman, L. A. Laimins, P. B. Sigler, Crystal structure at 1.7 A of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature* **359**, 505–512 (1992).
11. Y. Kim, J. H. Geiger, S. Hahn, P. B. Sigler, Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520 (1993).
12. J. L. Kim, D. B. Nikolov, S. K. Burley, Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520–527 (1993).
13. Y. Zhao, D. Granas, G. D. Stormo, Inferring binding energies from selected binding sites. *PLoS Comput. Biol.* **5**, e1000590 (2009).
14. A. Jolma *et al.*, Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Res.* **20**, 861–873 (2010).
15. M. Slattery *et al.*, Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* **147**, 1270–1282 (2011).
16. M. F. Berger *et al.*, Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.* **24**, 1429–1435 (2006).
17. C. L. Warren *et al.*, Defining the sequence-recognition profile of DNA-binding molecules. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 867–872 (2006).
18. R. Gordân *et al.*, Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep.* **3**, 1093–1104 (2013).
19. R. Nutiu *et al.*, Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nat. Biotechnol.* **29**, 659–664 (2011).
20. D. D. Le *et al.*, Comprehensive, high-resolution binding energy landscapes reveal context dependencies of transcription factor binding. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E3702–E3711 (2018).
21. A. Isakova *et al.*, SMiLE-seq identifies binding motifs of single and dimeric transcription factors. *Nat. Methods* **14**, 316–322 (2017).
22. G. D. Stormo, DNA binding sites: Representation and discovery. *Bioinformatics* **16**, 16–23 (2000).
23. G. D. Stormo, Y. Zhao, Determining the specificity of protein-DNA interactions. *Nat. Rev. Genet.* **11**, 751–760 (2010).
24. M. T. Weirauch *et al.*, Evaluation of methods for modeling transcription factor sequence specificity. *Nat. Biotechnol.* **31**, 126–134 (2013).
25. T. Zhou *et al.*, Quantitative modeling of transcription factor binding specificities using DNA shape. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 4654–4659 (2015).
26. B. Alipanahi, A. Delong, M. T. Weirauch, B. J. Frey, Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* **33**, 831–838 (2015).
27. D. Schübeler, Function and information content of DNA methylation. *Nature* **517**, 321–326 (2015).
28. D. Golovenko *et al.*, New insights into the role of DNA shape on its recognition by p53 proteins. *Structure* **26**, 1237–1250.e6 (2018).
29. S. Hoshika *et al.*, Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science* **363**, 884–887 (2019).
30. A. Afek *et al.*, DNA mismatches reveal conformational penalties in protein-DNA recognition. *Nature* **587**, 291–296 (2020).
31. M. A. M. Reijns *et al.*, Lagging-strand replication shapes the mutational landscape of the genome. *Nature* **518**, 502–506 (2015).
32. R. Sabarinathan, L. Mularoni, J. Deu-Pons, A. Gonzalez-Perez, N. López-Bigas, Nucleotide excision repair is impaired by binding of transcription factors to DNA. *Nature* **532**, 264–267 (2016).
33. M. Pacesa *et al.*, Structural basis for Cas9 off-target activity. *Cell* **185**, 4067–4081.e21 (2022).
34. D. Wang *et al.*, MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. *Nucleic Acids Res.* **45**, 2396–2407 (2017).
35. X. Lei *et al.*, Crystal structure of apo MEF2B reveals new insights in DNA binding and cofactor interaction. *Biochemistry* **57**, 4047–4051 (2018).
36. J. F. Kribelbauer *et al.*, Quantitative analysis of the DNA methylation sensitivity of transcription factor complexes. *Cell Rep.* **19**, 2383–2395 (2017).
37. H. M. Berman *et al.*, The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
38. P. Brownlie *et al.*, The crystal structure of an intact human Max-DNA complex: New insights into mechanisms of transcriptional control. *Structure* **5**, 509–520 (1997).
39. C. Rastogi *et al.*, Accurate and sensitive quantification of protein-DNA binding affinity. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E3692–E3701 (2018).
40. N. M. Luscombe, R. A. Laskowski, J. M. Thornton, Amino acid-base interactions: A three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.* **29**, 2860–2874 (2001).
41. A. C. Dantas Machado *et al.*, Landscape of DNA binding signatures of myocyte enhancer factor-2B reveals a unique interplay of base and shape readout. *Nucleic Acids Res.* **48**, 8529–8544 (2020).
42. A. Han *et al.*, Sequence-specific recruitment of transcriptional co-repressor Cabin1 by myocyte enhancer factor-2. *Nature* **422**, 730–734 (2003).
43. S. A. Coulocheri, D. G. Pigis, K. A. Papavassiliou, A. G. Papavassiliou, Hydrogen bonds in protein-DNA complexes: Where geometry meets plasticity. *Biochimie* **89**, 1291–1303 (2007).
44. Y. Liu, X. Zhang, R. M. Blumenthal, X. Cheng, A common mode of recognition for methylated CpG. *Trends Biochem. Sci.* **38**, 177–183 (2013).
45. W. S. El-Deiry, S. E. Kern, J. A. Pietenpol, K. W. Kinzler, B. Vogelstein, Definition of a consensus binding site for p53. *Nat. Genet.* **1**, 45–49 (1992).
46. Y. Yin *et al.*, Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, eaaj2239 (2017).
47. X. Liu, E. R. Weikum, D. Tilo, C. Vinson, E. A. Ortlund, Structural basis for glucocorticoid receptor recognition of both unmodified and methylated binding sites, precursors of a modern recognition element. *Nucleic Acids Res.* **49**, 8923–8933 (2021).
48. Z. Zuo, B. Roy, Y. K. Chang, D. Granas, G. D. Stormo, Measuring quantitative effects of methylation on transcription factor-DNA binding affinity. *Sci. Adv.* **3**, eaao1799 (2017).
49. R. Chatterjee *et al.*, High-resolution genome-wide DNA methylation maps of mouse primary female dermal fibroblasts and keratinocytes. *Epigenetics chromatin* **7**, 35 (2014).
50. H. Zhu, G. Wang, J. Qian, Transcription factors as readers and effectors of DNA methylation. *Nat. Rev. Genet.* **17**, 551–565 (2016).
51. J. Yang *et al.*, Structural basis for effects of CpA modifications on C/EBPβ binding of DNA. *Nucleic Acids Res.* **47**, 1774–1785 (2019).
52. K. Hoogsteen, Crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* **16**, 907–916 (1963).
53. E.-A. Raiber, R. Hardisty, P. van Delft, S. Balasubramanian, Mapping and elucidating the function of modified bases in DNA. *Nat. Rev. Chem.* **1**, 0069 (2017).

USC Michelson Center for Convergent Bioscience