
Supplementary information

Perspectives on ENCODE

In the format provided by the authors and unedited

Supplementary Information

Perspectives on ENCODE

ENCODE Perspective authors

The ENCODE Project Consortium[#], Michael P. Snyder^{1,2}, Thomas R. Gingeras³, Jill E. Moore⁴, Zhiping Weng^{4,5,6}, Mark B. Gerstein⁷, Bing Ren^{8,9}, Ross C. Hardison¹⁰, John A. Stamatoyannopoulos^{11,12,13}, Brenton R. Graveley¹⁴, Elise A. Feingold¹⁵, Michael J. Pazin¹⁵, Michael Pagan¹⁵, Daniel A. Gilchrist¹⁵, Benjamin C. Hitz¹, J. Michael Cherry¹, Bradley E. Bernstein¹⁶, Eric M. Mendenhall^{17,18}, Daniel R. Zerbino¹⁹, Adam Frankish¹⁹, Paul Flicek¹⁹, Richard M. Myers¹⁸

The ENCODE Project Consortium

The Broad Institute of Harvard and MIT (data production and analysis)

Charles B. Epstein²⁰, Noam Shores²⁰, Robbyn Issner²⁰, Shawn Gillespie²¹, Dylan Rausch²¹, Joseph Raymond²⁰, Shanna Hsu²⁰, Danielle Tenen²⁰, Oren Ram²⁰, Alon Goren²⁰, Russell Ryan²¹, Mariateresa Fulciniti²², David Hendrickson²⁰, Jonathan Scheiman²³, Birgit Knoechel^{22,24}, David Kelley²⁵, Samantha Beik²⁰, Yotam Drier²¹, Carles Boix²⁶, Wouter Meuleman¹¹, Pouya Kheradpour²⁶, Nina Farrell²⁰, Meital Hatan²⁰, David Wine²⁰, Mia C. Uziel²⁰, Kristin G. Ardlie²⁰, Michael Mannstadt²¹, Nikhil Munshi²², Miguel Rivera²¹, Alex Meissner²⁷, Manolis Kellis^{20,26}, John Rinn²⁸, Bradley E. Bernstein¹⁶

Cold Spring Harbor, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology and Universitat Pompeu Fabra (data production and analysis)

Carrie A. Davis³, Alexander Dobin³, Alessandra Breschi²⁹, Sarah Djebali^{29,30}, Chris Zaleski³, Dmitri D. Pervouchine^{29,31}, Anna Vlasova²⁹, Jorg Drenkow³², Julien Lagarde²⁹, Rory Johnson^{29,33}, Barbara Uszczynska-Ratajczak^{29,34}, Alexandra Scavelli³, Cassidy Danyko³, Lei Hoon See³, Roderic Guigó²⁹, Thomas R. Gingeras³

UConn Health, UCSD, Massachusetts Institute of Technology, and Institut de Recherches Cliniques de Montréal (IRCM) (data production and analysis)

Cassandra Bazile³⁵, Steven M. Blue³⁶, Eric L. Van Nostrand³⁶, Louis Philip Benoit Bouvrette^{37,38,39}, Daniel Dominguez³⁵, Peter Freese⁴⁰, Jia-Yu Chen⁴¹, Neal A.L. Cody^{37,38,39}, Gabriel A. Pratt³⁶, Michael O. Duff¹⁴, Sara Olson¹⁴, Xiaofeng Wang^{37,38,39}, Keri Elkins³⁶, Balaji Sundararaman³⁶, Xintao Wei¹⁴, Chelsea Anne Gelboin-Burkhart³⁶, Rui Xiao⁴¹, Abigail Hochman³⁵, Lijun Zhan¹⁴, Nicole J. Lambert³⁵, Hairi Li⁴¹, Thai B. Nguyen³⁶, Tsultrim Palden³⁵, Ines Rabano³⁶, Shashank Sathe³⁶, Rebecca Stanton³⁶, Amanda Su³⁵, Ruth Wang³⁶, Brian A. Yee³⁶, Xiang-Dong Fu⁴¹, Eric Lécuyer^{37,38,39}, Christopher B. Burge³⁵, Gene W. Yeo³⁶, Brenton R. Graveley¹⁴

HudsonAlpha Institute for Biotechnology, California Institute of Technology, The Pennsylvania State University, National Human Genome Research Institute, University of Alabama in Huntsville, Duke University, University of California Irvine (data production and analysis)

Mark Mackiewicz¹⁸, Florencia Pauli-Behn¹⁸, E. Christopher Partridge¹⁸, Daniel Savic⁴², Brian Roberts¹⁸, Kimberly M. Newberry¹⁸, Laurel A. Brandsmeier¹⁸, Sarah K. Meadows¹⁸, Rosy Nguyen¹⁸, Amy R. Nesmith¹⁸, Dianna E. Moore¹⁸, Christopher L. Messer¹⁸, Megan McEown¹⁸, Rachel C. Evans¹⁸, J Scott Newberry¹⁸, Collin White¹⁸, Shawn Levy¹⁸, Barbara Wold⁴³, Brian A. Williams⁴³, Diane E. Trout⁴³, Gilberto DeSalvo⁴³, Kenneth McCue⁴³, Peng He⁴³, Katherine Fisher-Aylor⁴³, Sean A. Upchurch⁴³, Henry Amrhein⁴³, Georgi K. Marinov⁴³, Jost Vielmetter⁴³, Anthony Kirilusha⁴³, Igor Antoshechkin⁴³, Ross C. Hardison¹⁰, Cheryl A. Keller¹⁰, Belinda M. Giardine¹⁰, Maria Long¹⁰, David M. Bodine⁴⁴, Elisabeth F. Heuston⁴⁴, Stacie M. Anderson⁴⁴, Eric M. Mendenhall^{17,18}, Surya B. Chhetri^{17,18}, Candice J. Coppola^{17,18}, Timothy E. Reddy^{45,46}, Anthony M. D'Ippolito⁴⁶, Christopher M. Vockley²⁰, Ali Mortazavi⁴⁷, Rabi Murad⁴⁷, Weihua Zeng⁴⁷, Camden Jansen⁴⁷, Ricardo N. Ramirez⁴⁷, Nicole El-Ali⁴⁷, Richard M. Myers¹⁸

University of California, San Diego, Salk Institute for Biological Studies, Lawrence Berkeley National Laboratory, UC San Diego, Howard Hughes Medical Institute (data production and analysis)

David U. Gorkin^{8,9}, Yupeng He⁴⁸, Iros Barozzi⁴⁹, Andre Wildberg⁵⁰, Jennifer A. Akiyama⁴⁹, Rosa G. Castanon⁴⁸, Sora Chee⁸, Huaming Chen⁴⁸, Bo Ding⁵⁰, Yoko Fukuda-Yuzawa⁴⁹, Tyler H. Garvin⁴⁹, Manoj Hariharan⁴⁸, Anne Harrington⁴⁹, Hui Huang⁸, Momoe Kato⁴⁹, Samantha Kuan⁸, Ah Young Lee⁸, Elizabeth Lee⁴⁹, Bin Li⁸, Nan Li⁵⁰, Brandon J. Mannion⁴⁹, Joseph R. Nery⁴⁸, Vu Ngo⁵⁰, Tung Nguyen⁵⁰, Quan Pham⁴⁹, Catherine S. Novak⁴⁹, Ingrid Plajzer-Frick⁴⁹, Sebastian Preissl⁸, Yin Shen^{8,51}, Mengchi Wang⁵⁰, Tao Wang⁵⁰, John W. Whitaker⁵⁰, Yanxiao Zhang⁸, Kai

Zhang⁵⁰, Yuan Zhao⁸, Yun Zhu⁵⁰, Feng Yue^{52,53}, Veena Afzal⁴⁹, Diane E. Dickel⁴⁹, Valentina Snetkova⁴⁹, Wei Wang⁵⁰, Axel Visel^{49,54,55}, Len A. Pennacchio^{49,54,56}, Joseph R. Ecker^{48,57}, Bing Ren^{8,9}

Stanford University, The University of Chicago, University of Southern California, University of Toronto, Yale University (data production and analysis)

Jessika Adrian¹, Trupti Kaul¹, Nicholas J. Addleman¹, Alan P. Boyle^{58,59}, Lulu Cao¹, Hassan Chaib¹, Songjie Chen¹, Catharine Eastman⁶⁰, Andrew Emili⁶¹, Peng Gong⁶², Fabian Grubert¹, Yu Guo⁶³, Hongbo Guo⁶¹, Nastaran Heidari^{1,64}, Cory Holgren⁶⁵, Nader Jameel⁶⁵, Lixia Jiang¹, Madhura Kadaba⁶⁵, Maya Kasowski¹, Mary Kasparian⁶⁵, Yining Li¹, Jin Lian⁶², Yiing Lin⁶⁶, Shin Lin¹, Lijia Ma⁶⁵, Matthew G. Milton⁶⁵, Tejaswini Mishra¹, Jennifer Moran⁶⁵, Anil M. Narasimha¹, Xinghua Pan^{62,67,68}, Doug H. Phanstiel^{69,70}, Ernest Radovani⁶¹, Lucia Ramirez¹, Rozita Razavi⁷¹, Suhn K. Rhie⁶³, Denis N. Salins¹, Frank W. Schmitges⁷¹, Quan Shen^{62,72}, Minyi Shi¹, Teri Slifer¹, Damek V. Spacek¹, Rohith Srivas¹, Dave Steffan⁶⁵, Matt Szynekarek⁶⁵, Dave Toffey⁶⁵, Alec Victorsen⁶⁵, Nathaniel K. Watson¹, Heather N. Witt⁶³, Xinqiong Yang¹, Jie Zhai¹, Jialing Zhang⁶², Guoqing Zhong⁶¹, Sherman M. Weissman⁶², Jack F. Greenblatt^{61,71}, Timothy R. Hughes^{61,73}, Peggy J. Farnham⁶³, Kevin P. White⁷⁴, Michael P. Snyder^{1,2}

Altius Institute for Biomedical Sciences, University of Washington, Fred Hutchinson Cancer Research Center, University of Massachusetts Medical School, Howard Hughes Medical Institute (data production and analysis)

John A. Stamatoyannopoulos^{11,12,13}, Rajinder Kaul^{11,13}, Jessica Halow¹¹, Richard Sandstrom¹¹, Michael Buckley¹¹, Jeff Vierstra¹¹, Wouter Meuleman¹¹, Eric Haugen¹¹, Shane Neph¹¹, Andrew Nishida¹¹, Alex Reynolds¹¹, Eric Rynes¹¹, Audra Johnson¹¹, Jemma Nelson¹¹, Alister P. W. Funnell¹¹, Vivek Nandakumar¹¹, Kyle Siebenthal¹¹, Hao Wang¹¹, Yongqi Yan¹¹, Reyes Acosta¹¹, Kristen Lee¹¹, Ericka Otterman¹¹, Benjamin Van Biber¹¹, Mineo Iwata¹¹, Tanya Kutayavin¹¹, Sandra Stehling-Sun¹¹, Robert E. Welikson¹¹, Andres Castillo¹¹, Grigorios Georgolopoulos¹¹, Sean Ibarrientos¹¹, Fidencio Jun Neri¹¹, Anthony Shafer¹¹, Shinny Vong¹¹, Daniel Bates¹¹, Morgan Diegel¹¹, Douglass Dunn¹¹, John Lazar¹¹, Daniel R. Chee¹¹, George Stamatoyannopoulos⁷⁵, Patrick Navas¹¹, M. A. Bender⁷⁶, Mark T. Groudine⁷⁶, Rachel Byron⁷⁶, Ye Zhan⁷⁷, Hakan Ozadam⁷⁷, Bryan R. Lajoie⁷⁷, Job Dekker⁷⁷

Data Coordination Center at Stanford (data coordination center)

J. Michael Cherry¹, Benjamin C. Hitz¹, Cricket A. Sloan¹, Ulugbek K. Baymuradov¹, Esther T.

Chan¹, Timothy R. Dreszer¹, Idan Gabdank¹, Jason A. Hilton¹, Aditi K. Narayanan¹, Kathrina C. Onate¹, J. Seth Strattan¹, Forrest Y. Tanaka¹

University of Massachusetts Medical School, Yale University, Stanford University, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology and Universitat Pompeu Fabra, University of Washington, Dana-Farber Cancer Institute, Harvard School of Public Health, Massachusetts Institute of Technology (data analysis center)

Jill E. Moore⁴, Michael J. Purcaro⁴, Henry E. Pratt⁴, Gregory R. Andrews⁴, Tyler Borrman⁴, Jason A. Brooks⁴, Hao Chen^{4,6}, Shaimae I. Elhajjajy⁴, Kaili Fan⁴, Kevin Fortier⁴, Mingshi Gao⁴, Jack Huey⁴, Eugenio Mattei⁴, Nishigandha N. Phalke⁴, Thomas M. Reimonn⁴, Shuo Shan⁴, Junko Tsuji⁴, Arjan G. van der Velde^{4,6}, Taylor Young⁴, Xiao-Ou Zhang⁴, Tianxiong Yu⁵, Peng Zhang⁵, Lingling Cai⁵, Xiangrui Li⁵, Ying Li⁵, Zhijie Wei⁵, Shuyu Hou⁵, Aiping Lu⁵, Yan Liu⁵, Yu Fu^{4,78}, Shaliu Fu⁵, Qin Wang⁵, Joel Rozowsky⁷, Jing Zhang⁷, Koon-Kiu Yan⁷, Mengting Gu⁷, Donghoon Lee⁷, Shaoke Lou⁷, Anurag Sethi⁷, Sushant Kumar⁷, Timur Galeev⁷, Arif Harmanci⁷, Gamze Gursoy⁷, Jason Liu⁷, Jinrui Xu⁷, Fabio C.P. Navarro⁷, Nathan Boley¹, Daniel Sunwook Kim¹, Jin Wook Lee¹, Chuan Sheng Foo⁷⁹, Oana Ursu¹, Sarah Djebali^{29,30}, Dmitri D. Pervouchine^{29,31}, Anna Vlasova²⁹, Beatrice Borsari²⁹, Maxwell W. Libbrecht⁸⁰, Galip Gürkan Yardımci¹², Peng Jiang^{81,82}, Clifford Meyer^{81,82}, Chongzhi Zang^{81,82,83}, Qian Qin^{23,84}, Mingxiang Teng⁸⁵, Jose Davila-Velderrain²⁶, Yue Li²⁶, Zhizhuo Zhang²⁶, Dianbo Liu²⁶, Carles Boix²⁶, Yongjin Park²⁶, Yaping Liu²⁶, Lei Hou²⁶, Manolis Kellis^{20,26}, X. Shirley Liu^{81,82}, William S. Noble¹², Roderic Guigó²⁹, Anshul Kundaje¹, Mark B. Gerstein⁷, Zhiping Weng^{4,5,6}

Massachusetts Institute of Technology (computational algorithm development)

Amira A. Barkal⁸⁶, Budhaditya Banerjee⁸⁷, Matthew D. Edwards²⁶, David K. Gifford²⁶, Yuchun Guo²⁶, Tatsunori B. Hashimoto²⁶, Tommi Jaakkola²⁶, Charles W. O'Donnell²⁶, Nisha Rajagopal⁸⁷, Richard I. Sherwood⁸⁷, Sharanya Srinivasan⁸⁷, Tahin Syed²⁶, Haoyang Zeng²⁶

University of Wisconsin–Madison, University of Nebraska-Lincoln, The Ohio State University, University of Wisconsin School of Medicine and Public Health (computational algorithm development)

Ye Zheng⁸⁸, Peng Liu⁸⁹, Sunyoung Shin⁹⁰, Rene Welch⁸⁹, Jurijs Nazarovs⁸⁸, Qi Zhang⁹¹, Dongjun Chung⁹², Emery H. Bresnick⁹³, Colin N. Dewey⁸⁹, Sunduz Keles^{88,89}

Icahn School of Medicine at Mount Sinai, Brigham and Women's Hospital and Harvard Medical School, Memorial Sloan Kettering Cancer Center (computational algorithm development)

James E. Hayes⁹⁴, Gosia Trynka⁹⁵, Alvaro Gonzalez⁹⁶, Harm-Jan Westra⁸⁷, Manu Setty⁹⁶, Maria Gutierrez-Arcelus⁸⁷, Yuheng Lu⁹⁶, Alexander R. Perez⁹⁶, Yuri Pritykin⁹⁶, Mark Carty⁹⁶, Christina S. Leslie⁹⁶, Soumya Raychaudhuri⁸⁷, Robert J. Klein⁹⁴

Stanford (computational algorithm development)

Anand Bhaskar¹, Yang I. Li¹, Graham McVicker⁴⁸, Eilon Sharon¹, Anil Raj¹, Jonathan K. Pritchard¹

University of California, Los Angeles (computational algorithm development)

Yun-Hua E. Hsiao⁹⁷, Giovanni Quinones-Valdez⁹⁷, Yi-Wen Yang⁹⁷, Xinshu Xiao⁹⁷

Johns Hopkins University (data analysis)

Michael A. Beer^{98,99}

Pennsylvania State University/Northwestern University (data production and analysis)

Yanli Wang⁵³, Hongbo Yang⁵³, Tingting Liu⁵³, Lijun Zhang⁵³, Jie Xu⁵³, Bo Zhang⁵³, Feng Yue^{52,53}

European Bioinformatics Institute (EMBL-EBI), Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology and Universitat Pompeu Fabra, ELIXIR Hub, King's College London, Guy's Hospital, Massachusetts Institute of Technology, Spanish National Cancer Research Centre (CNIO), University of Bern, University of California, Santa Cruz, University of Lausanne, Wellcome Sanger Institute, Yale University (gene annotation)

Bronwen Aken¹⁹, Joel Armstrong¹⁰⁰, Matthew Astley⁹⁵, If H.A. Barnes¹⁹, Daniel Barrell¹⁹, Gemma Barson⁹⁵, Ruth Bennett¹⁹, Andrew Berry¹⁹, Alexandra Bignell¹⁹, Jacqueline Chrast¹⁰¹, Declan Clarke⁷, Claire Davidson¹⁹, Alden Deran¹⁰⁰, Gloria Despacio-Reyes⁹⁵, Mark Diekhans¹⁰⁰, Sarah Donaldson¹⁹, Iakes Ezkurdia¹⁰², Anne-Maud Ferreira¹⁰¹, Stephen Fitzgerald⁹⁵, Carlos Garcia Giron¹⁹, Jose M. Gonzalez¹⁹, Michael Gray⁹⁵, Ed Griffiths⁹⁵, Matthew Hardy¹⁹, Toby Hunt¹⁹, Rory Johnson^{29,33}, Irwin Jungreis^{20,26}, Michael Kay¹⁹, Julien Lagarde²⁹, Jane Loveland¹⁹, Deepa Manthravadi⁹⁵, Osagie Izuogu¹⁹, Fergal J. Martin¹⁹, Steve Miller⁹⁵, Jonathan M. Mudge¹⁹, Eva Maria Novoa²⁶, Baikang Pei⁷, Dmitri D. Pervouchine^{29,31}, Jose M. Rodriguez¹⁰³, Christoph

Schlaffner⁹⁵, Cristina Sisu^{7,104}, Marie-Marthe Suner¹⁹, Michael L. Tress¹⁰³, Barbara Uszczynska-Ratajczak^{29,34}, Jesus Vazquez¹⁰², Hendrik Weisser⁹⁵, Maxim Wolf²⁶, James Wright⁹⁵, Tim J. Hubbard¹⁰⁵, Jennifer L. Harrow¹⁰⁶, Alfonso Valencia¹⁰³, Federico Abascal⁹⁵, Manolis Kellis^{20,26}, Paul Flicek¹⁹, Benedict Paten¹⁰⁰, Jyoti S. Choudhary¹⁰⁷, Mark B. Gerstein⁷, Alexandre Reymond¹⁰¹, Roderic Guigó²⁹, Adam Frankish¹⁹

Florida State University and University of Georgia (data production and analysis)

Juan Carlos Rivera-Mulia^{108,109}, Takayo Sasaki¹⁰⁸, Vishnu Dileep¹⁰⁸, Jared Zimmerman¹⁰⁸, Michael J. Kulik¹¹⁰, Stephen Dalton¹¹¹, David M. Gilbert¹⁰⁸

European Bioinformatics Institute (EMBL-EBI) (genome annotation)

Emily H. Perry¹⁹, Daniel R. Zerbino¹⁹, Paul Flicek¹⁹

The Broad Institute of Harvard and MIT, Gift of Life Donor Program, American Society for Radiation Oncology, National Cancer Institute (NCI), Leidos Biomedical, Inc., National Disease Research Interchange (NDRI), National Human Genome Research Institute (NHGRI) (tissue sample preparation)

Kristin G. Ardlie²⁰, Richard D. Hasz¹¹², Judith C. Keen¹¹³, Helen M. Moore¹¹⁴, Anna Smith¹¹⁵, Jeffrey A. Thomas¹¹⁶, Simona Volpi¹⁵

National Human Genome Research Institute (Project Management)

Xiao-Qiao Zhou¹⁵, Hannah Naughton¹⁵, Julie Coursen¹⁵, Samuel H. Moore¹⁵, Preetha Nandi¹⁵, Omar Al Jammal¹⁵, Yekaterina Vaydylevich¹⁵, Peter Good¹¹⁷, Jeffery A. Schloss¹⁵, Briana Nuñez¹⁵, Michael Pagan¹⁵, Eileen Cahill¹⁵, Daniel A. Gilchrist¹⁵, Michael J. Pazin¹⁵, Elise A. Feingold¹⁵

(The role of the NHGRI Project Management Group in the preparation of this paper was limited to coordination and scientific management of the ENCODE consortium.)

Affiliations

- 1) Department of Genetics, School of Medicine, Stanford University, Palo Alto, California 94305, USA.
- 2) Cardiovascular Institute, Stanford School of Medicine, Stanford CA 94305
- 3) Cold Spring Harbor Laboratory, Functional Genomics, 1 Bungtown Road, Cold Spring Harbor, New York 11742, USA

- 4) University of Massachusetts Medical School, Program in Bioinformatics and Integrative Biology, 368 Plantation St., Worcester, MA 01605, USA
- 5) Department of Thoracic Surgery, Clinical Translational Research Center, Shanghai Pulmonary Hospital, The School of Life Sciences and Technology, Tongji University, Shanghai 200092, China
- 6) Bioinformatics Program, Boston University, Boston, MA 02215, USA
- 7) Yale University, New Haven, Connecticut 06520-8047, USA
- 8) Ludwig Institute for Cancer Research, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0653
- 9) Center for Epigenomics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0653
- 10) Department of Biochemistry and Molecular Biology, The Pennsylvania State University, 304 Wartik Laboratory, University Park PA 16802
- 11) Altius Institute for Biomedical Sciences, 2211 Elliott Avenue, Suite 410, Seattle, WA, 98121
- 12) Department of Genome Sciences, University of Washington, 3720 15th Ave NE Seattle, Washington 98195-5065, USA.
- 13) Department of Medicine, RR-512, Health Sciences Building, University of Washington, Box 356420, 1959 NE Pacific Street, Seattle, WA 98195-6420
- 14) Department of Genetics and Genome Sciences, Institute for Systems Genomics, UConn Health, Farmington, CT 06030, USA
- 15) National Human Genome Research Institute, National Institutes of Health, 6700B Rockledge Drive, Bethesda, MD 20817
- 16) Broad Institute and Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114
- 17) Biological Sciences, University of Alabama in Huntsville, 301 Sparkman Drive, Huntsville AL 35899
- 18) HudsonAlpha Institute for Biotechnology, 601 Genome Way, Huntsville, AL 35806
- 19) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom
- 20) The Broad Institute of Harvard and MIT, 415 Main St, Cambridge, MA 02142
- 21) MGH, 55 Fruit St, Boston, MA 02114
- 22) Dana-Farber Cancer Institute, 450 Brookline Ave, Boston, MA 02215
- 23) Harvard Medical School, 25 Shattuck St, Boston, MA 02115
- 24) Boston Children's Hospital, 300 Longwood Ave, Boston, MA 02115

- 25) Harvard University, Massachusetts Hall, Cambridge, MA 02138
- 26) Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139
- 27) Max Planck Institute for Molecular Genetics, Department of Genome Regulation, Ihnestr. 63-73, Berlin 14195 Germany
- 28) University of Colorado Boulder, Boulder CO 80301
- 29) Bioinformatics and Genomics Program, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology and Universitat Pompeu Fabra, Dr. Aiguader 88, 08003 Barcelona, Catalonia, Spain
- 30) IRSD, Université de Toulouse, INSERM, INRA, ENVT, UPS, U1220, CHU Purpan, CS60039, 31024, Toulouse, France
- 31) Skolkovo Institute for Science and Technology, Bolshoy Boulevard 30, bld. 1 Moscow 121205, Russia
- 32) Cold Spring Harbor Laboratory, 500 Sunnyside Blvd, Woodbury NY 11797, USA
- 33) Department of Clinical Research, University of Bern, Murtenstrasse 35, 3010 Bern, Switzerland.
- 34) International Institute of Molecular and Cell Biology, Ks. Trojdena 4 02-109 Warsaw Poland
- 35) Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA
- 36) Department of Cellular and Molecular Medicine, Institute for Genomic Medicine, Stem Cell Program, Sanford Consortium for Regenerative Medicine, University of California, San Diego, La Jolla, CA 92037, USA
- 37) Département de Biochimie et Médecine Moléculaire, Université de Montréal, Montréal, QC H2W 1R7, Canada
- 38) Division of Experimental Medicine, McGill University, QC H2W 1R7, Canada
- 39) Institut de Recherches Cliniques de Montréal (IRCM), Montréal, QC H2W 1R7, Canada
- 40) Program in Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA
- 41) Department of Cellular and Molecular Medicine, Institute of Genomic Medicine, University of California at San Diego, San Diego, California 92093, USA
- 42) Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN 38105, USA
- 43) Division of Biology 156-29, 1200 E. California Blvd, California Institute of Technology, Pasadena CA 91125

- 44) National Human Genome Research Institute, National Institutes of Health, Building 49, Room 4A04, 40 Convent Drive, Bethesda, MD 20892
- 45) Department of Biostatistics and Bioinformatics, Duke University, Durham, NC 27708, USA
- 46) Center for Genomic and Computational Biology, Duke University, Durham, NC 27708, USA
- 47) 2218 Biological Sciences III, University of California Irvine, Irvine CA 92697-2300
- 48) Salk Institute for Biological Studies, 10010 N Torrey Pines Rd, La Jolla, CA 92037
- 49) Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
- 50) Department of Chemistry and Biochemistry, Department of Cellular and Molecular Medicine, UC San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0359
- 51) Institute for Human Genetics, Department of Neurology, University of California, San Francisco, San Francisco, California 94143, USA
- 52) Department of Biochemistry and Molecular Genetics, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA.
- 53) Penn State Health Milton S. Hershey Medical Center, 500 University Dr., Hershey, PA 17033, USA
- 54) U.S. Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
- 55) School of Natural Sciences, University of California, Merced, Merced, California, USA
- 56) Comparative Biochemistry Program, University of California, Berkeley, CA 94720, USA
- 57) Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 N Torrey Pines Rd, La Jolla, CA 92037
- 58) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA.
- 59) Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA.
- 60) Department of Molecular and Cellular Physiology, School of Medicine, Stanford University, Palo Alto, California 94305, USA.
- 61) Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada.
- 62) Department of Genetics, School of Medicine, Yale University, New Haven, Connecticut 06520, USA.
- 63) Department of Biochemistry and Molecular Medicine, Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, California 90089, USA.

- 64) Department of Radiation Oncology, School of Medicine, Stanford University, Palo Alto, California 94305, USA
- 65) Department of Human Genetics, Institute for Genomics and Systems Biology, The University of Chicago, Chicago, Illinois 60637, USA.
- 66) Division of General Surgery, Section of Transplant Surgery, School of Medicine, Washington University, St. Louis, Missouri 63110, USA.
- 67) Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, Guangzhou, China
- 68) Guangdong Provincial Key Laboratory of Single Cell Technology and Application, Guangzhou 510515, Guangdong, China
- 69) Department of Cell Biology & Physiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599
- 70) Thurston Arthritis Research Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599
- 71) Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada.
- 72) School of Medicine, Jiangsu University, Zhenjiang, Jiangsu 212013, China
- 73) Department of Molecular Genetics, Donnelly Centre, University of Toronto, Toronto, Ontario M5S 3E1, Canada
- 74) Tempus Labs, 600 W Chicago Ave. Ste 510, Chicago, IL 60654
- 75) Department of Medicine, University of Washington, K-240 Health Sciences Building, Box 357720, Seattle, WA 98195
- 76) Fred Hutchinson Cancer Research Center, 1100 Fairview ave. N, Seattle, WA 98109
- 77) HHMI and Program in Systems Biology, University of Massachusetts Medical School, 368 Plantation Street, Albert Sherman Center, Worcester, MA 01605
- 78) University of Massachusetts Amherst, 270 Stockbridge Rd, Amherst, MA 01003, USA
- 79) Institute for Infocomm Research, 1 Fusionopolis Way, #21-01 Connexis (South Tower), Singapore 138632, Singapore
- 80) Simon Fraser University, ASB 9971-8888 University Drive. Burnaby, British Columbia V5A 1S6, Canada
- 81) Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA 02215
- 82) Department of Data Sciences, Dana-Farber Cancer Institute and Harvard T.H. Chan School of Public Health, Boston, MA 02215

- 83) Center for Public Health Genomics, University of Virginia, PO Box 800717, Charlottesville, VA 22908-0717, USA
- 84) Molecular Pathology Unit & Cancer Center, 55 Fruit St Boston MA 02114
- 85) Department of Biostatistics and Bioinformatics, Moffitt Cancer Center, 12902 Magnolia Drive, Tampa, FL 33612
- 86) Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, 265 Campus Drive, Stanford, CA 94305
- 87) Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 77 Ave Louis Pasteur, Boston, MA 02215
- 88) Department of Statistics, Medical Sciences Center, 1300 University Ave, Madison, WI 53706 USA
- 89) Department of Biostatistics and Medical Informatics, 425 Henry Mall 2130 Genetics/Biotechnology Center, Madison, Wisconsin 53706, USA
- 90) Department of Mathematical Sciences at University of Texas at Dallas, 800 W Campbell Rd, FO 3.608 Richardson, TX 75080
- 91) Department of Statistics, University of Nebraska-Lincoln, 354C Hardin Hall North Wing, Lincoln, NE 68583-0963 USA
- 92) The Ohio State University, Department of Biomedical Informatics, 1800 Cannon Drive, 250 Lincoln Tower, Columbus, OH 43210
- 93) Department of Cell and Regenerative Biology, UW-Madison Blood Research Program, Carbone Cancer Center, University of Wisconsin School of Medicine and Public Health, University of Wisconsin, Madison, Wisconsin 53705, USA
- 94) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, 1 Gustave L Levy Place, New York, NY 10029
- 95) Wellcome Sanger Institute, Cambridge, CB10 1 SA, UK
- 96) Program in Computational Biology, Memorial Sloan Kettering Cancer Center, 1275 York Ave, New York, NY 10065
- 97) 610 Charles E. Young Drive S, Terasaki Life Sciences Building, Room 2000E, Los Angeles, CA 90095-1570
- 98) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland, USA
- 99) Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, USA
- 100) University of California, Santa Cruz, California 95064, USA

- 101) Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland
- 102) Centro Nacional de Investigaciones Cardiovasculares (CNIC) and CIBER de Enfermedades Cardiovasculares (CIBERCV), Madrid, Spain
- 103) Spanish National Cancer Research Centre (CNIO), E-28029 Madrid, Spain
- 104) Brunel University London, London, UB8 3PH, UK
- 105) King's College London, Guy's Hospital, Great Maze Pond, London, SE1 9RT, UK
- 106) ELIXIR Hub, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK
- 107) Institute of Cancer Research, Chester Betty Labs, South Kensington, London, SW3 6JB, UK
- 108) Department of Biological Science, 319 Stadium Drive, Florida State University, Tallahassee, FL 32306, USA
- 109) Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota Medical School, Minneapolis, MN 55455
- 110) Center for Vaccines and Immunology University of Georgia, CVI Room 1506, 501 DW Brooks Drive, Athens, GA 30602, USA
- 111) Center for Molecular Medicine and Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA
- 112) Gift of Life Donor Program, 401 North 3rd Street, Philadelphia, PA 19123
- 113) American Society for Radiation Oncology, 251 18th Street South, 8th Floor, Arlington, VA 22202
- 114) National Cancer Institute, National Institutes of Health, 9609 Medical Center Drive, Bethesda, MD 20892
- 115) Leidos Biomedical, Inc., 8560 Progress Drive, Frederick, MD 21701
- 116) National Disease Research Interchange (NDRI), 8 Penn Center, 15th Floor, 1628 JFK Boulevard, Philadelphia, PA 19103
- 117) 4407 Puller Dr, Kensington, MD 20895

Contact author: mpsnyder@stanford.edu

Supplementary Note 1 Useful URLs

ENCODE Portal

- a) ENCODE Portal: <https://www.encodeproject.org>

Using ENCODE Data

- b) ENCODE Data use policy: <https://www.encodeproject.org/about/data-use-policy/>
- c) ENCODE Tutorials and handouts: <https://www.encodeproject.org/tutorials/>

Accessing ENCODE data

- d) ENCODE encyclopedia: <https://www.encodeproject.org/data/annotations/>
- e) ENCODE SCREEN: <http://screen.encodeproject.org>
- f) ENCODE Registry of candidate regulatory elements:
https://www.encodeproject.org/matrix/?type=Annotation&annotation_type=candidate+regulatory+elements&files.file_type=bed+bed3%2B
- g) Factorbook of transcription factors assayed by ENCODE: <http://factorbook.org>
- h) REST API: <https://www.encodeproject.org/help/rest-api/>
- i) All ENCODE data:
<https://www.encodeproject.org/matrix/?type=Experiment&award.project=ENCODE>
 - 1. all ENCODE mouse data:
https://www.encodeproject.org/matrix/?type=Experiment&award.project=ENCODE&replicates.library.biosample.donor.organism.scientific_name=Mus+musculus
 - 2. all ENCODE human data:
https://www.encodeproject.org/matrix/?type=Experiment&award.project=ENCODE&replicates.library.biosample.donor.organism.scientific_name=Homo+sapiens
 - 3. all modENCODE/modERN data:
<https://www.encodeproject.org/matrix/?type=Experiment&award.project=modENCODE&award.project=modERN>
 - i. all worm data:
https://www.encodeproject.org/matrix/?type=Experiment&award.project=modENCODE&award.project=modERN&replicates.library.biosample.donor.organism.scientific_name=Caenorhabditis+elegans
 - ii. all fly data:
<https://www.encodeproject.org/matrix/?type=Experiment&award.project=modENCODE&award.project=modERN>
- 4. ENCODE EN-TEEx samples:
https://www.encodeproject.org/matrix/?searchTerm=gtex&type=Experiment&status=released&replicates.library.biosample.biosample_type=tissue&y.limit=
- 5. ENCODE 3 mouse matrix samples:
https://www.encodeproject.org/matrix/?searchTerm=ENCODE+3&type=Experiment&award.rfa=ENCODE+3&replicates.library.biosample.donor.organism.scientific_name=Mus+musculus&biosample_type=tissue&organ_slims=brain&organ_slims=heart&organ_slims=liver&organ_slims=limb&organ_slims=epithelium&organ_slims=lung&organ_slims=stomach&organ_slims=kidney&organ_slims=intestine&organ_slims=embryo

Examples of how ENCODE data are used

- j) Community and ENCODE consortium publications: <https://www.encodeproject.org/publications/>

How ENCODE data are generated and processed

- k) ENCODE Experimental guidelines/protocols and data standards/quality metrics: <https://www.encodeproject.org/data-standards/>
- l) ENCODE data processing pipelines: <https://www.encodeproject.org/pipelines/>
- m) Software tools: <https://www.encodeproject.org/software/>
- n) Ontologies used by ENCODE: <https://www.encodeproject.org/help/getting-started/#Ontologies>
 - 1. experimental assay types (OBI: http://obi-ontology.org/page/Main_Page)
 - 2. cell lines (EFO: <http://www.ebi.ac.uk/efo/>)
 - 3. primary cells (CL: <http://cellontology.org/>)
 - 4. tissues (UBERON: <http://uberon.org/>)
- o) Antibody characterization: <https://www.encodeproject.org/search/?type=AntibodyLot>
- p) ENCODE Help Desk: encode-help@lists.stanford.edu

ENCODE Participants

- q) ENCODE 3 participants: <https://www.genome.gov/27568353/encode-phase-iii-participants-and-projects/>
- r) ENCODE 4 participants: <https://www.genome.gov/26525220/encode-participants-and-projects/>

ENCODE is a member of these organizations

- s) International Human Epigenome Consortium (IHEC): <http://ihec-epigenomes.org/welcome/>
- t) Global Alliance for Genomics and Health (GA4GH): <https://www.ga4gh.org>

ENCODE Tissues and Cell Lines: Current and Planned

- u) <https://www.encodeproject.org/proposed-biosamples/>

Related Projects

- v) NIH Roadmap Epigenomics Program (REMC): <https://commonfund.nih.gov/epigenomics/index>
- w) International Human Epigenome Consortium (IHEC): <http://ihec-epigenomes.org>
- x) Canadian Epigenetics, Environment and Health Research Consortium (CEEHRC) Network: <http://www.epigenomes.ca>
- y) Blueprint: <http://www.blueprint-epigenome.eu>
- z) 4D Nucleome Program (4DN): <https://commonfund.nih.gov/4Dnucleome/index>
- aa) The Cancer Genome Atlas (TCGA): <https://cancergenome.nih.gov>
- bb) Genotype and Tissue Expression Project (GTEx): <https://commonfund.nih.gov/GTEx/index>
- cc) PsychENCODE: https://www.nimhgenetics.org/available_data/psychencode/
- dd) Functional Annotation of the Mammalian Genomes (FANTOM) <https://fantom.gsc.riken.jp/>
- ee) The Human Cell Atlas (HCS) <https://www.broadinstitute.org/research-highlights-human-cell-atlas>
- ff) The Human Biomolecular Atlas Program (HuBMAP) <https://commonfund.nih.gov/hubmap>
- gg) Functional Annotation of Animal Genomes (FAANG): <http://www.faang.org>