CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Bard College

Bard College
Bard Digital Commons

Senior Projects Spring 2017
Bard Undergraduate Senior Projects

Spring 2017

Denatured: Emergent realities of encyclopedic DNA elements

Amelia Leeya Goldstein
Bard College, ag2933@bard.edu

Follow this and additional works at: https://digitalcommons.bard.edu/senproj_s2017


Part of the Other Genetics and Genomics Commons, Other Sociology Commons, Science and Technology Studies Commons, and the Theory, Knowledge and Science Commons

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 License.

Recommended Citation

Goldstein, Amelia Leeya, "Denatured: Emergent realities of encyclopedic DNA elements" (2017). *Senior Projects Spring 2017*. 356.
https://digitalcommons.bard.edu/senproj_s2017/356

This Open Access work is protected by copyright and/or related rights. It has been provided to you by Bard College's Stevenson Library with permission from the rights-holder(s). You are free to use this work in any way that is permitted by the copyright and related rights. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself. For more information, please contact digitalcommons@bard.edu.

Bard

Denatured:
Emergent realities of encyclopedic DNA elements

A Senior Project Submitted to
The Division of Social Studies
of Bard College

by

Amelia Leeya Hyatt Goldstein

Annandale-on-Hudson, NY
May 2017

Acknowledgements

Thank-you to my Sociology family that helped me through this. Laura, thank you for your dedication to me and for your patience and endurance with me. Peter, thank you for being a huge source of support and encouragement for the past three years, and for helping me out when I got lost in this project, so many times. Yuval, thank you for encouraging and supporting me. Allison, thank you for teaching me so much, for being my advisor these past two years, and for being on my board. To my fellow sociology seniors, for being in this with me.

Thank you to all of the other professors that were a part of this with me: Diana thank you so much for being there for me in this project, for encouraging me, and for believing in me. Sophia, thank you for introducing me to my new friend Bruno Latour. Yuka, thank you for your relentless support and encouragement. Heidi, your help in shaping this project was invaluable, even if none of it is recognizable. Brooke, thank you for introducing me to ENCODE, for all of your support and encouragement, which let me believe that I was capable of talking about science in this way. Thank you to Gabriel, Arseny, and Mike, for reminding me to be both hopeful and nihilistic in the future of evolutionary biology.

Thank you to my parents and grandparents for helping me through college and making me do this. Thank you to my cousins, aunts, and uncles as well for supporting me relentlessly.

Thank you to my closest friends, nothing I do would be possible without you. Thank you for the sproj breaks, the love, and the music.

Thank you to Carl, for *everything*, and to Betsy and Lily, for making the library a welcoming and encouraging place for me! Thank you Jeremy for all of your help and support. Thank you to everyone who gave me love and support on long library days and cold carrel nights.

Thank you so, so much to everyone who proofread and peer reviewed this paper. You have been my angels.

Thank you to all the people who wrote alongside me, who cheered me on, who gave me hope when I lost it. Thank you to everyone who loved me and put up with me during this grueling process and all of my griping and panicking, to my CEAP family, to my Harvesting Justice crew, and to you, my reader.

Preface: A Birth Story.

My mother told me that before I was a person, but after I had DNA, she tried to undergo amniocentesis, to see me, or to see my DNA, they were hardly distinct yet. I held my little hand out and grabbed the needle in utero. I would not let go. They stopped the procedure and never tried again.

My mother says it was a miracle that I was born without In-Vitro Fertilization.
Nowadays, genetic screening is often a mandatory part of the In-Vitro Fertilization procedure.

Why did I reach for the needle? Was it because of something I had in me already? How much of us is inside of us before we are born? Some say all of us, but it seems that many geneticists would not agree with that anymore. Maybe I reached for the needle to keep from having too much inside of me before being born. Maybe it was there, but I wanted it to stay invisible.

When I think about getting my DNA tested now, I think about how I will be granted genetic citizenship to a country I have never visited. I think about the cancer in my family, my father's glaucoma, filling out family histories at the doctor's office. I think about the adipose tissues I incessantly reminded myself as a teenager were artifacts of my genes, not my choices. I think about how my eyes don't match my parents' and how I convinced myself when I was young that because of this I must have been adopted; and that I must have been adopted from someone very important for my parents to have kept it secret so long. I think about how it is an evolutionary anomaly to inherit a condition that makes it impossible, or at least miraculous, to reproduce without the use of infertility technologies. I think about sitting in a hospital waiting room at 15 and swearing I would never pass these genes on to anyone else.

This project has become a frame analysis of the Encyclopedia of DNA Elements, to see where genomic information is stopped or limited. I thought at first that I was writing this because I wanted to frame my own DNA, but maybe I just do not want my DNA to frame me. At this point, maybe all I want is for my DNA not to tell me where I end and where disease begins.

Abstract:

The Human Genome Project was the center of much controversy in the 1990's, as creating a map of the human genome drew into question the boundaries between nature and nurture, or science and society. Fifteen years have now passed since the Human Genome Project's completion, and the new paradigm of genetics is no longer governed by a strict nature/ nurture dualism. This project looks at one of the Human Genome Project's successors: the Encyclopedia of DNA Elements (ENCODE) project, which has created new boundaries and limitations in this new phase of genetic thinking. Using a frame analysis and Actor-Network Theory approach to follow how ENCODE has formed and reformed over the years, this project traces the ENCODE project as a new way of translating genetic code from the cell to the world around it, and ultimately back into the cell. Throughout these processes, the ENCODE project brings into question the meaning of human, creates a platform for viewing the genome as a moldable substance, and ultimately presents itself as the end of human disease.

"Biology is a political discourse, one in which we should engage at every level of practice- technically, semiotically, morally, economically, institutionally. And besides all that, biology is a source of intense intellectual, emotional, social, and physical pleasure. Nothing like that should be given up lightly- or approached only in a scolding mode."

-Donna Haraway, *Modest_Witness@Second_Millennium.FemaleMan©_Meets_ OncoMouse™,* p. 105

"Technology is sociology extended by other means"

-Bruno Latour, *Aramis,* p. 210

CONTENTS

CHAPTER 1. INTRODUCTION.

Introducing ENCODE

When the United States National Institute of Health and Department of Energy launched the Human Genome[1] Project in the 1990's, it became one of the largest international scientific consortiums to date. The National Institute of Health founded a new branch to carry out their component to this project: The National Human Genome Research Institute.[2] This Maryland-based institute would take over the National Institute of Health in overseeing not only the Human Genome Project, but whatever projects came next. Sociologists have referred to this time as the beginning of "The New Genetics," a movement in genetic science's ever-expanding territory to encapsulate complex traits including diseases, behaviors, and personality traits.[3] Geneticists understood this trend as a shift from using biology to understand how genetics works, to instead using genetics to understand how biology, or other topics such as behavior or race, work.[4] Accordingly, sociologists asked questions about just how far genetics will go, or where the boundaries are that intercept the movement of biology.[5,6,7] The Human Genome Project officially finished its task of writing the code for the entire human genome in 2001.

---

[1] The word *genome* refers to all of the genetic material present in an organism.

[2] National Human Genome Research Institute Division of Genome Sciences. "About NHGRI: A Brief History and Timeline" National Human Genome Research Institute. Accessed January 2017.

[3]Conrad, Peter, and Jonathan Gabe. *Sociological Perspectives on the New Genetics*. Malden, MA: Blackwell Publishers, 1999.

[4]Siddhartha Mukherjee. *The gene: An intimate history*. Simon and Schuster, 2016.

[5] Anne Kerr, Sarah Cunningham-Burley and Amanda Amos. *The new genetics: professional's discursive boundaries.* Sociological Review, 42, 2, 1997.

[6]Ann Juanita Morning. *The Nature of Race: How Scientists Think and Teach about Human Difference*. Berkeley: U of California, 2011.

[7]Aaron Panofsky. *Misbehaving Science: Controversy and the Development of Behavior Genetics*. Chicago: University of Chicago, 2014.

In 2002, the National Human Genome Research Institute held a workshop to discuss a proposal to form a new research consortium, a pilot project to test and compare the identification of functional sequences in the human genome.[8] The description of this workshop, archived by the National Human Genome Research Institute, was that it would synthesize the many approaches that had been developed to characterize "functional elements" in genomic DNA. Unlike the Human Genome Project, which was about the pure code of DNA, this new project would elucidate the three-dimensional structure of DNA and the complexities of DNA and RNA[9] transcription. This project saw a problem in the lack of unity in the approaches for identifying genomic elements, and in the lack of comprehensive methods for testing the roles of genomic elements. The National Human Genome Research Institute wanted groups working in these areas to work together and to communicate with one another.

To achieve this goal, the National Human Genome Research Institute proposed the formation of a new consortium. It would be open to any scientists from academic, governmental, and private sector projects that wanted to participate, and it would start with a unified goal: to select a limited region of the human genome, and get participation and insight from a variety of participants in order to analyze and interpret that section of the genome. The National Human Genome Research Institute hoped that ultimately, the functional elements of the entire human genome would be made clear, but that the original pilot project would start small, with only 1% of the human genome.[10]

---

[8] For a history of functional, versus structural, genomics, look to:
Evelyn Fox Keller *The Century of the Gene.* Cambridge: Harvard UP. 2000.

[9] RNA are molecules which mirror DNA through a process called transcription, one of the many steps through which DNA leads to proteins.

[10] "Workshop Summary" *National Human Genome Research Institute,* Last reviewed: March 19, 2012. Accessed January 2017.

At this workshop, presenters shared their perspectives on computational and experimental approaches that would confirm functional elements in the genomes of both model organisms and humans, and outlined a proposal for a pilot project to "exhaustively determine all functional elements" in the selected 1% of the human genome.[11] The consortium would be a community that agreed to focus on this small portion of the human genome, and was willing to share their results openly and immediately to one another, and contribute in conversation with other participants. Participants at the workshop decided that there needed to be an ontology, a set of shared vocabulary[12] developed for the consortium to share. This was the beginning of the formation of ENCODE: the Encyclopedia of DNA Elements, a research project and accompanying consortium hosted by the National Human Genome Research Institute.

In 2003, the National Human Genome Research Institute held a meeting to discuss the launch of the ENCODE project and consortium. The National Human Genome Research Institute shared the first two of many ENCODE grants and funding opportunities, which allowed it to set the stage for what ENCODE would be and provide the parameters for participation. The ENCODE project would build from the Human Genome Project by creating data, tools, and analyses that put the long string of bases (letters that make up DNA) into context. The following image, published recently by the National Human Genome Research Institute, shows how the National Human Genome Research Institute imagines the role of ENCODE as a project which builds from the Human Genome Project's long code:

---

[11] "Workshop Summary"

[12] Many ENCODE participants, including ENCODE Data Coordination Center's Primary Investigator, J Michael Cherry, are involved in the Gene Ontology Consortium, which can be found at www.geneontology.org, which was referenced at this meeting as a model ontology.

Figure 1.1. 'Genome in 3-D' image from the National Human Genome Research Institute in February 2011[13]

The Human Genome Project wrote the two-dimensional code of A's, C's, T's, and G's, or bases, that make up human DNA. The ENCODE project takes this information and folds it up, exploring the dynamics that shape it and allow it to function in a three-dimensional world. "At its core," a recent news release from the National Institute of Health stated, "ENCODE is about enabling the scientific community to make discoveries by using basic science approaches to understand genomes at the most fundamental level."[14]

Participants at the 2003 ENCODE launch meeting came from a variety of institutions, and became the original founders of ENCODE. Each participant represented its own institution, but they all came together to consider participation in the ENCODE project, and to provide a critical perspective on what it would look like. 15 participants were from the NIH or NHGRI, and the rest represented various research institutes, universities, or biomedical companies.

---

[13] The four images in this chapter are from the NHGRI, which allows its images to be free of use as long as they are cited to the NHGRI. Bard College has no policy currently on image use in senior projects, aside from the standard plagiarism policy which requires citations for all sources.

[14] "NIH to expand critical catalog for genomics research" *National Institute of Health*, February 2, 2017. https://www.nih.gov/news-events/news-releases/nih-expand-critical-catalog-genomics-research

Participation was majority white, English-speaking, and extended internationally only as far as Canada, Spain, and England. Each institution represented was either a research institution or a biotechnology company, all with different contributions to offer to this consortium. By bringing this diverse group together, ENCODE was formed at a crossroads of various categories of institutions, each coming from different incentives, be them the pursuit of knowledge, health, or capital, all under the auspices of the National Human Genome Research Institute.

Since 2003, the ENCODE Consortium's goals have expanded significantly. The first phase of ENCODE, which sought to understand 1% of the human genome, was completed in 2007, and phase two of the ENCODE project began. To accommodate this change in scale, the National Human Genome Research Institute released a grant in 2006 for the development of a Data Coordination Center for the ENCODE project, which is now an entity hosted by the Stanford University research laboratory, and a grant to form a Data Analysis Center, which is now hosted at the University of Massachusetts Medical School in Worcester, Massachusetts. In 2007, the ENCODE project expanded to include model species genomes through the modENCODE and a modERN consortia. To facilitate this growth, the GENCODE consortium was formed, which produces reference gene annotations for ENCODE. Its aim is to "annotate all evidence-based gene features on the human genome" so as to fill some of the gaps between how much of the genome is possible to sequence and how much is possible to decipher and understand.[15] This expansion in 2006 is sometimes referred to as ENCODE's 'scale-up' project.[16]

---

[15] Jennifer Harrow et al. "GENCODE: the reference human genome annotation for The ENCODE Project." *Genome research 22.9* (2012): 1760-1774.

[16]"The ENCODE Project: ENCyclopedia of DNA Elements" National Human Genome Research Institute. Last Updated, February 2017. https://www.genome.gov/10005107/encode-project/#al-4

Throughout the decade and a half since the ENCODE project was first proposed, ENCODE has formed many new partnerships and affiliations, both as a project to create an encyclopedia and as a project to connect different scientific research projects. These partnerships include: *Nature*, the prestigious academic journal; various branches of the National Institute of Health, including the Roadmap Epigenomics Mapping Center, which was launched in 2007 to perform an ENCODE-adjacent map of the *epigenome*;[17] and bioinformatics projects at the University of California Santa Cruz and the Wellcome Trust Sanger Institute in the United Kingdom. Participation is open to any research institution, public, private, or international. Data production centers are listed on the National Human Genome Research Institute website,[18] and come from various universities and research institutes across the United States.

Primary Investigators from various laboratories work on different research components that fit together to form the complete ENCODE project. For example, Bradley Bernstein's laboratory is focused on a particular type of histone (a protein involved in the shaping and regulation of the DNA molecule) mapping, and Michael Snyder's group identifies transcription factor binding sites (specific sections of DNA to which enzymes involved in the transcription of DNA latch).[19] In addition to production centers, there are a few other types of institutions which

---

[17] The *epigenome* refers to DNA and RNA regulatory elements that are changed throughout an organism's lifespan through a process called methylation, in which methyl groups are added onto the DNA molecule, turning genes 'on' or 'off.' *Epigenetics*, subsequently, is the study of how genes vary in expression, rather than changes in genetic codes.

[18] National Human Genome Research Institute- ENCODE Project. "ENCODE Participants and Projects." National Human Genome Research Institute. Accessed January 2017.

[19] National Human Genome Research Institute- ENCODE Project. "ENCODE Participants and Projects."

make up the ENCODE consortium,[20,21] whose goals include developing models to predict regulatory elements, determining the function of DNA elements, annotation of gene features through GENCODE, investigating relationships between gene timing and expression patterns,[22] developing regulatory data, and developing computational methods of analysis.

The new affiliated consortiums that have developed since ENCODE's expansion include GENCODE, modENCODE, modERN, Roadmap and the Genomics of Gene Regulation. The model species projects modENCODE and modERN look at the genomes of *Drosophila* (fruit flies) and *C. Elegans* (worms). Roadmap and the Genomics of Gene Regulation are project from the National Institute of Health that have information hosted on the ENCODE genome browser, but are a separate research projects. ENCODE has also begun to display its data on a genome browser provided by a London-based project called *Ensembl,* which launched in 1999, with a similar mission to ENCODE's of making data from the Human Genome Project more useful, connected, and accessible to scientists.[23,24.]

---

[20] These include (as of January, 2017): various research groups that work on technology development from Massachusetts Institute of Technology, Washington University at St. Louis, the University of Southern California, University of Washington, Memorial Sloan-Kettering Cancer Center, the University of North Carolina at Chapel Hill, University of Michigan, Broad Institute of MIT and Harvard, University of Washington, Advanced RNA Technologies LLC, and Harvard School of Public Health; and participants from John's Hopkins, Wellcome Trust Sanger Institute, Florida State University, and University of Southern California's Keck School of Medicine.

[21] National Human Genome Research Institute- ENCODE Project. "ENCODE Participants and Projects." National Human Genome Research Institute. Accessed January 2017.

[22] These refer to patterns in how genes come to form traits, and the timing of this process, through effects such as circadian rhythms that regulate genetic responses on set schedules.

[23] Ensembl. "ENCODE data in Ensembl" *e!ensembl.* published December 2016, accessed February 2017. http://useast.ensembl.org/info/website/tutorials/encode.html

[24] Ensembl. "About the Ensembl Project" *e!ensembl*. Ensembl release 2017. http://www.ensembl.org/info/about/index.html

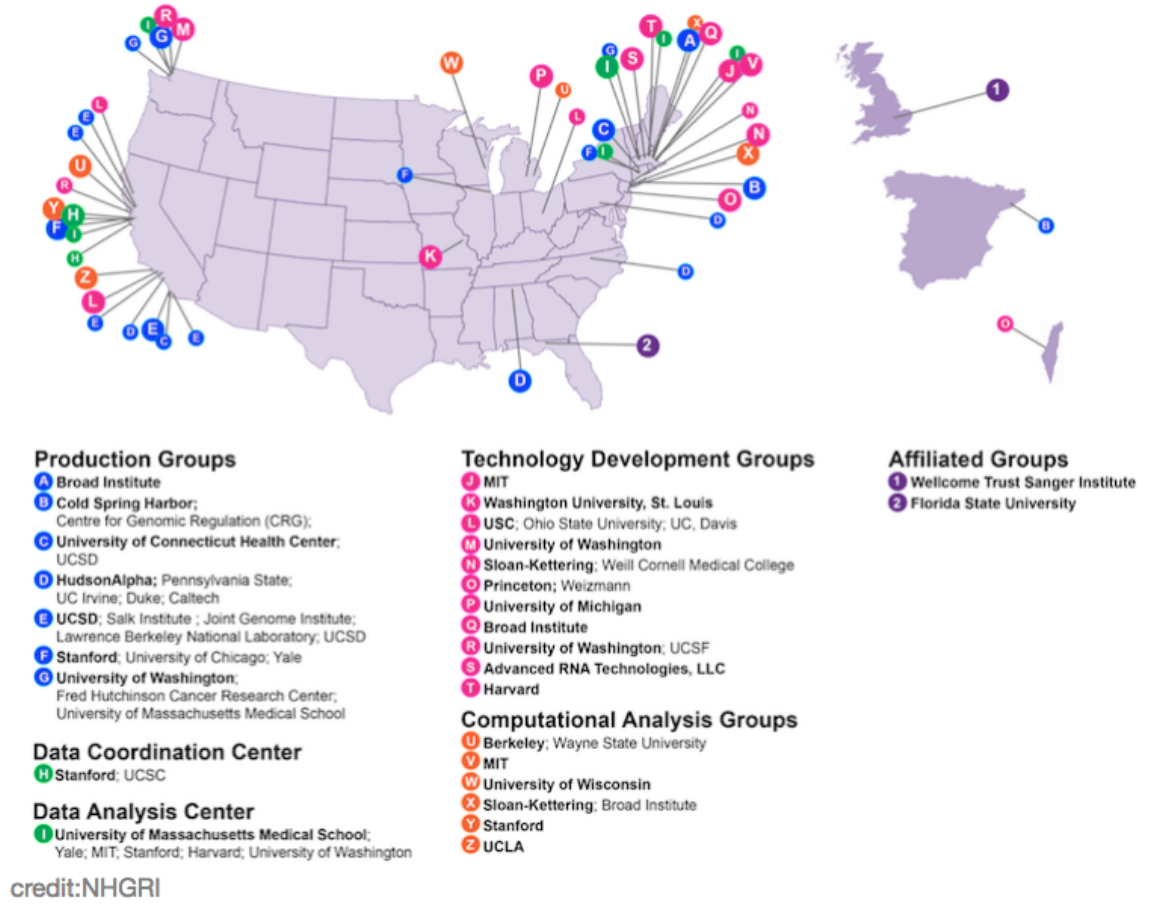


Figure 1.2. A map of ENCODE participation. In addition to the institutes in the United States, 1 points to the Wellcome Trust Sanger Institute in England, 'O' points to the Weizmann institute in Israel and 'B' points to the Centre for Genomic Regulation in Barcelona.[25]

Leadership

The directors of ENCODE, from the time of its launch to present day, are Elise Feingold, Ph.D, Mike Pazin, Ph.D, and Daniel Gilchrist, Ph.D. The Data Coordination Center, which consists of a team of data wranglers and wrangling assistants, is led by J. Michael Cherry, PhD. Wranglers allow data to be handled and organized, so that data production can move forward. If ENCODE were written as a many-volume, printed encyclopedia, the wranglers are the librarians, who make sure that it is ordered and intact, assign call numbers, and direct patrons to the

[25] Data Coordination Center. "Project Overview." *Encode: the Encyclopedia of DNA Elements*. Accessed December 2016. https://www.encodeproject.org/about/contributors/

volumes they seek. They are the search engines, the editors, and the programmers. With the increasingly large scale in ENCODE's bioinformatics, data wranglers play an important role in framing and presenting data. They are curators, and stand in for publishers in a project that is moving too fast for a standard peer review publishing process. They take on the role of translation, in bringing data from the laboratory, or bringing computational data from the computer, to an integrated matrix of ENCODE data.

Data production is an open process, and although the National Human Genome Research Institute does fund data production for the ENCODE project, funding may come from external projects. Data production uses certain tools of analysis: assays are run which achieve the ENCODE mission of elucidating and validating functional DNA elements. Genome-Wide Association Studies are used to reveal which regions should be looked at further, or 'candidate regions,' as well as to identify single point mutations through the identification of single nucleotide polymorphisms (or SNPs, pronounced 'snip's, these are the main point of variation between genomes of the same species, and are used to identify tiny mutations, or differences, which can in translation refer to anything from physical appearance to genetic 'disease').[26]

In an ENCODE presentation by director Mike Pazin, the following image is used to describe the ENCODE consortium structure:

---

[26]ENCODE Project Consortium. "An integrated encyclopedia of DNA elements in the human genome." *Nature* 489, no. 7414 (2012): 57-74.

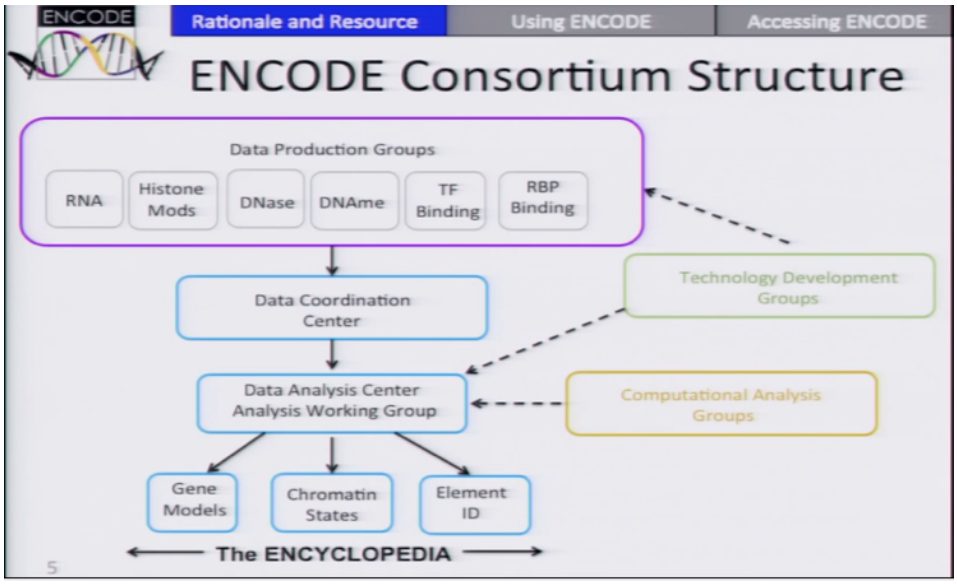


Figure 1. 3. ENCODE Consortium Structure Slide from 2015 presentation. Taken from a presentation by Mike Pazin.[27]

This figure shows that the main categories of ENCODE data production are DNA assays that work with RNA isolation and sequencing (RNA are long codes of nucleotides, the same substance that makes up DNA, that mirror DNA), the modifications of histones, sequencing of DNAse (an enzyme that breaks down the DNA molecule into smaller units), DNAme (the methylation of DNA, or *epigenetics*[28]), or the mapping of transcription factor binding sites in either DNA or RNA. There are various tools developed by or used by ENCODE-funded researchers that enable it to locate or identify these parts of DNA.[29] Additionally, many ENCODE researchers run computational assays to predict specific genetic elements.

This figure also depicts the continuum in which ENCODE data is circulated. It starts in the production centers and enters the ENCODE Data Coordination Center, where data is

---

[27] National Human Genome Research Institute. "Using ENCODE Data to Interpret Disease-associated Genetic Variation- Mike Pazin," Youtube Video. Published August 21, 2019: 3:33.

[28] See footnote 17.

[29] These are all written about in the following article published by the ENCODE consortium:
ENCODE Project Consortium. "An integrated encyclopedia of DNA elements in the human genome." *Nature* 489, no. 7414 (2012): 57-74.

wrangled and sorted, and then is organized and presented by data 'analysts,' who bring data into the ENCODE encyclopedia, which currently takes the form of a matrix on encodeproject.org. This data is then publicized and shared in the form of publications, conversations, presentations, and social media posts. Through these channels, data produced by ENCODE researchers, through ENCODE funding, enters the the clinic, the classroom, and other elements of the social world. The ENCODE Data Coordination Center even joined Twitter in 2011, and the ENCODE project followed shortly thereafter in 2012.

Why Does ENCODE Matter to Sociology?

The ENCODE project was founded in order to continue researching genetics with the resources established by the Human Genome Project. The Human Genome Project developed a long string of letters- three billion A's, C's, T's, and G's to be specific- and stopped there. Its existence caused a large stir of controversy from not only scientists and bioethicists, but sociologists, political scientists, and more,[30,31] who looked at how the Human Genome Project, by creating the entire code of the human, drew into question the meaning of race and ethnicity,[32,33] and how DNA was becoming a social icon.[34] Soon, *The New Genetics* became a topic of many sociological analyses, looking at the implications of the Human Genome Project

---

[30]Kevles, Daniel J. "The Code of Codes: Scientific and Social Issues in the Human Genome Project. By Daniel J. Kevles and Leroy E. Hood. Cambridge, MA: Harvard UP, 1992.

[31] Paul Rabinow. "Artificiality and Enlightenment: From Sociobiology to Biosociality." *Anthropologies of Modernity: Foucault, Governmentality, and Life Politics*. By Jonathan Xavier Inda. Malden, MA: Blackwell Pub., 2005. 181-193.

[32] Dorothy E. Roberts. *Fatal Invention: How Science, Politics, and Big Business Re-create Race in the Twenty-first Century*. New York: New, 2011.

[33] Timothy F. Murphy. "The Genome Project and the Meaning of Difference." *Justice and the Human Genome Project.* By Timothy F. Murphy and Marc Lappel. Berkeley: U of California, 1994. 1-13.

[34] Nelkin and Lindee. *The DNA Mystique.*

and other genetic technologies and discoveries beyond designated scientific spaces.[35,36] Other social scientists have looked at specific genetic traits that question the boundaries between science and sociology, such as the 'gay gene,'[37] genetic causes of mental illness, and the genetics of chemical sensitivity.[38] Some put genetics into a context of obstetrics, looking at how prenatal genetic screening has created a process called the "new eugenics."[39] Many follow a dichotomy of nature versus nurture, bringing in sociological perspectives as at odds with the scientific findings to write sociological critiques of genetic sciences,[40,41] or practices of genetic essentialism.[42,43]

The biological paradigm which pits sociology against biology has been called into question by a rise in genetics perspectives which emphasize plasticity and epigenetics.[44] Epigenetics, the study of regulatory effects from the environment on the genome, includes social contextual factors in understanding how the genome works. As a result, biological projects increasingly look at these regulatory effects: less on the code itself or the trait itself, but on the mediators in between.

---

35 Conrad and Gabe. *Sociological Perspectives on the New Genetics*

36 Kerr et al. *The New Genetics.*

37 Deborah Lynn Steinberg *Genes and the Bioimaginary: Science, Spectacle, Culture*: Taylor & Francis, 2015.

38 Shostak, Sara, Peter Conrad, and Allan V. Horwitz. "Sequencing and Its Consequences: Path Dependence and the Relationships between Genetics and Medicalization 1." *American Journal of Sociology* 114, no. S1 (2008): S287-S316.

39 Carole H. Browner and Nancy Ann Press. "The Normalization of Prenatal Diagnostic Screening." Conceiving the New World Order: The Global Politics of Reproduction. By Faye. D. Ginsburg and Rayna Rapp. Berkeley: University of Califonia Press, 1995. 307-21.

40 Nelkin and Lindee. *The DNA Mystique.*

41 Morning, Ann Juanita. *The Nature of Race: How Scientists Think and Teach about Human Difference*. Berkeley: U of California, 2011.

42Kerr et al. *The New Genetics.*

43 Panofsky. *Misbehaving Science.*

44 Laland et al. "Does evolutionary theory need a rethink?" Nature 514, no 7521 (2014): 161.

This is what my analysis seeks to provide: a look into this process through which the code produced by the Human Genome Project is translated into the traits that make up human life, and a look into the after-effects of public critiques of the Human Genome Project, by exploring its successor. This project seeks to provide a foundation, by asking how ENCODE, since its conception, has formed and reformed, using tools from Erving Goffman's *Frame Analysis* to trace the frames that have been articulated around the ENCOE project.

The second chapter of this essay will use Goffman's *Frame Analysis* to set a foundation of how the ENCODE project sets its own limits and boundaries on how it mediates information from the body to laboratories and computers, then discuss the changes in subjects of ENCODE research from the ENCODE 'pilot project' launch in 2003 to the full ENCODE project launch in 2007. When ENCODE launched in 2003, it was only focused on humans and on 1% of the human genome, but as it entered phase two and expanded to include the whole genome in 2007, it also launched modENCODE and modERN, two projects that accompanied and were deeply affiliated with ENCODE, but were focused on model species. Although ENCODE still remained a project to characterize the human genome, this change indicates ambiguity in the frame of the "human genome" and in the boundary between human and nonhuman. What else changed in this moment of growth and frame expansion? How were the new boundaries of human versus nonhuman drawn, if they still existed?

The third chapter looks in particular at the framing of the ENCODE Data Analysis Center and Data Coordination Center, and how DNA moved through the frames established in Chapter Two as ENCODE entered phase three in 2012. ENCODE from 2003-2012 presented all of its data on platforms called genome browsers that were not specific to the ENCODE project, but in

2012, the ENCODE Data Coordination Center established its own website: encodeproject.org. ENCODE values of immediate data release were strengthened by ENCODE's newfound regulation of its own data release and presentation. This chapter explores what is enabled or disabled by the emergent ability for ENCODE entities to frame and display their own data, through the emergence of the actual ENCODE encyclopedia produced by the ENCODE Data Analysis and Data Coordination Center.

In the fourth chapter, I ask what happens when ENCODE, as a unified project, enters into conversations, controversies, and social media. In what ways is the ENCODE project uncontained? What are the new territories of ENCODE's future? In many ways, ENCODE is a concept or a topic, a subject of debate among scientists. When ENCODE enters the public eye, one of the biggest critiques it faces comes from politically charged geneticist Dan Graur, who published a controversial article in the journal of Genome Biology and Evolution on the use of the word "function" in publications by the ENCODE consortium. In this chapter, I ask what happens when the ENCODE project enters new territories such as social media, which shows me how the ENCODE project looks like to those outside of it, and elucidates some of the project's newest and outermost frames.

CHAPTER 2. ENCODE CASTS ITS ANCHOR.

Introduction

In 2003, the National Human Genome Research Institute held a meeting to announce and discuss the launch of a new project, years in the making, called ENCODE. This project, the founders described, was the answer to people's questions of 'what's next?' after the Human Genome Project came to an end.[45] Many speakers presented, most of whom held leadership in the National Human Genome Research Institute or who were founders to the ENCODE project. Francis Collins, a leader of the Human Genome Project, presented first. He explained that a meeting held the year prior, in which the ENCODE project was first publicly conceived, had endorsed a concept of focusing on 1% of the human genome at first, and that the ENCODE project would stick to that idea very strongly. This project, which would stick to 1% of the human genome, would be called the ENCODE pilot project. Collins explained that the ENCODE project would focus on *only* the human genome, and that "this is not a project that is focused on a model organism, it is focused on us, unless you want to call us a model organism."[46]

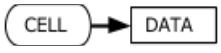
This statement established ENCODE's initial boundaries, limits, and parameters of research. ENCODE would be human-centered, but would not yet encapsulate the entire human. The boundaries of 1% and human centrism created a set of parameters and norms as to how to look at physical-form DNA from a cellular or molecular level, and translate it into data.[47] Anyone who wished to be funded by ENCODE grants, work with the ENCODE consortium, or

---

45 NIH Center for Information Technology. Encode Project Launch, Videocast (March 7 2003, Washington DC) https://videocast.nih.gov/summary.asp?Live=2306&bhcp=1: 00:01:00

46 NIH Center for Information Technology: 00:10:23

47 As illustrated in footer.



be cited and consulted as an ENCODE member would abide by these limitations. As years have passed since this meeting in 2003, however, these boundaries and limits have been pushed and challenged in many ways, by individuals, or by the entire ENCODE consortium. The purpose of this chapter is to draw on Erving Goffman's theory of Frame Analysis to understand these limits and examine how they transformed when ENCODE expanded in 2007.

Goffman first introduced the idea of the frame analysis in 1974 to explore how context changes the meaning of words by creating frames which encapsulate all of human interaction. He explained how cues, or *keys* as he called them, contribute to the meanings of different social interactions. Understanding what is and is not real, he argues, is not a strong enough tool for understanding what constitutes *reality.*[48] Instead, a frame analysis can render explicit these cues to provide social context, and thereby enable sociologists to characterize social interactions.

Much of Goffman's writing has been interpreted by social movement theorists to analyze how social movement organizations create frames,[49,50] but his work can also be understood as a tool for interpreting scientific behavior. He wrote that certain "intelligent agents" are given trust to understand the "natural world," but that even the most intelligent of agents will encounter certain "constraints,"[51] which the frame analysis can help to explain. He also wrote that "in our society, the very significant assumption is generally made that all events-without exception-can be contained and managed within the conventional system of beliefs. We tolerate the unexplained

---

[48] Erving Goffman, *Frame analysis: An essay on the organization of experience*. Harvard University Press, 1974: 247.

[49] Dawn McCaffrey and Jennifer Keys. "Competitive framing processes in the abortion debate: Polarization−vilification, frame saving, and frame debunking." *The Sociological Quarterly* 41, no. 1. 2000: 41-61.

[50] David A. Snow, E. Burke Rochford Jr., Steven K. Worden and Robert D. Benford. "Frame Alignment Processes, Micromobilization, and Movement Participation" American Sociological Review, Vol. 51, No. 4, August 1986 p 464-481.

[51] Goffman, *Frame analysis:* 1974: 23.

but not the inexplicable."[52] A frame analysis, he then explained, is a crucial element to making inexplicable social phenomena explicable, as well as to understand how others make inexplicable social phenomena explicable, and thus it can provide a home for sociological research in new domains or territories, outside of that which sociologists would otherwise have the tools to analyze. The frame analysis allows me to establish the norms and expectations of the ENCODE project, so as to then see where it transgresses its own boundaries, or where the boundaries themselves change.

Goffman primarily used the example of theatre to explain his concepts of frame analysis. In this example, the theoretical concept of *keying* can be understood as the public markers that designate what is within and outside of a play. For example, if somebody punches their friend, this may seem threatening and chaotic to other people in the room, but if somebody punches their friend on a stage, with an audience, in a costume, within the bounded time that the play takes place, it can be understood as an act of theatre, and seems perfectly safe to onlookers. Keys both determine and are determined by the practices of a given site. Keys are not stagnant, and therefore upkeying, downkeying, and rekeying can explain their movement. *Downkeying* is when the play becomes the theatergoer's reality, such as if the person who got punched was actually seriously injured, and left the stage to seek medical attention. *Upkeying* is when the reality becomes a play, such as when audience members begin to think they, too, are part of the play, maybe if they are called onto stage or addressed directly by an actor. *Rekeying* is when these small moments that make up an action's frame are shifted around, such as if the punch was in the script yesterday, but today the punch was taken out of the script, and the actor did it anyways.

---

[52] Goffman, *Frame analysis:* 1974: 247.

In applying the frame analysis to the ENCODE project, I use actor-network theory as a tool to make sense of ENCODE's development. Actor-Network Theory is a tool to understand science as a type of movement, and to not rely on *a priori* notions of science as one realm and the social as another, but instead to understand specific scientific actions as movements, associations, and assemblies. Actor Network-Theory calls for an equalization of all subjects of study and a flattening of the world which it describes. Actors, whether they be humans, animals, tools, platforms, organizations, or any number of things, can all be traced as they move through society in a mode that follows behaviors based on what is presented, instead of seeking to show gaps or silences in social behaviors. In this way, all actors are given equal agency, and a sociological analysis of *science* is just as possible as a sociology of *scientists,* which allows science to be held responsible for the way it moves and develops, rather than being seen as simply ontological and pure.[53]

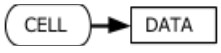
Often, when social scientists look at framing in the physical sciences, they take on the role of 'boundary-marking,' a concept introduced by sociologist of science Thomas Gieryn.[54,55] These studies focus on where science ends and the social world begins, mapping or characterizing lines between the two. The frame analysis, especially in combination with an Actor-Network Theory approach, differs from this in that it allows the social scientist to start with nothing, pick a site (the ENCODE genome), and work outwards, rather than assuming where a site is situated in the world. Context, then, will be built piece by piece, and be the end

---

[53] The concept of a pure science is taken from:
Pierre Bourdieu. "The Peculiar History of Scientific Reason." *Sociological Forum* no. 6.1, 1991: 3–26.

[54] Thomas Gieryn. *Cultural Boundaries of Science*. University of Chicago Press, 1999.

[55] Anne Kerr, Sarah Cunningham-Burley and Amanda Amos. *The new genetics: professional's discursive boundaries.* Sociological Review, 42, 2, 1997.



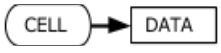
result of research rather than the starting-off point. Instead of finding DNA as at odds with society, I may see genomics as social processes, and look at what enables the social world of the genome to exist in ENCODE. While Gieryn enters a world of science as a sociologist equipped with an *a priori* notion of what science is versus what the social is, I allow ENCODE actors to set their own meanings for what is scientific versus what is social, or perhaps to drop this dualism altogether.

The hybrid of Actor-Network Theory and frame analysis allows me to understand ENCODE as a unique and specific project with its own ways of mediating the movement of DNA through society. Together, they allow me to build ENCODE's frames without *a priori* notions of what would be expected of this project, and thus allow me to look at this genomics project in a new light, to see where it bumps into other actors, and why it moves and grows in the ways it does. This mediation, or translation,[56,57] is a process with many intricacies and characteristics which this project should bring to light, with many implications for how a physical molecule can enter a cybernetic world, and eventually loop back into matter.

Historian of science Donna Haraway has used Actor-Network Theory to look at the Human Genome Project, particularly as it related to projects which work on mouse genomes, including OncoMouse™: a mouse with human breast cancer genes implanted in it. She says that to look at the world of technoscience ethnographically, we must pick a site, start at the inside, and work our way outward, allow metaphor and material to blend, and be open to new types of relationships or kinships. Thus, I begin with the translation of DNA from inside of the cell

[56] Translation studies is a common sort of Actor-Network Theory, focused on maintaining symmetry between two concepts. While this project is not formally a translational study, in following the translation of DNA from the cell to data, I am influenced by the Actor-Network Theorist Michael Callon's sociology of translation.

[57] Michel Callon. "Some elements of a sociology of translation: domestication of the scallops and the fishermen of St. Brieuc Bay." *The Sociological Review.* 32, no. S1, 1984: 196-233.



outward, and follow it further and further away from its cellular form. Haraway has said that if we want to understand the way the OncoMouse™ impacts humans, we need to enter the mouse and look outwards by looking at what enables OncoMouse™'s existence and what is enabled by OncoMouse™.[58] Instead of looking from the human genome's perspective outward, I may look at what has enabled or made possible the genome as it is understood by the ENCODE project or the way DNA is made possible translated into data. To do so, I must be open to a blend of the material and metaphor, and new types of relationships.

The rest of this chapter will be devoted to the keys provided in the initial 2003 ENCODE launch meeting, where ENCODE was described as a project to analyze 1% of the human genome, and the framing elements that have stemmed from, or challenged, that anchoring when ENCODE expanded in 2007. Along the way, I hope to characterize these frames using the tools provided by Goffman and those who have expanded on his theory, and Actor-Network Theory, as a research method for following the human genome through the ENCODE project.

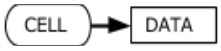
Anchoring ENCODE

The anchoring of activity, a concept from Goffman's *Frame Analysis*,[59] refers to the initial frame setting. In the example of theatre as a frame, anchoring moments would be when audience members buy tickets, when the playbill writes who is and is not in the cast, when the stage is set and the seats are set as two separate places. All of these are keys for the framing of theatre, but a specific type of key which allows audience members to initially establish that they

---

[58] Donna Haraway, "Mice into Wormholes," in Downey, Gary Lee, and Joseph Dumit, eds. *Cyborgs & citadels: anthropological interventions in emerging sciences and technologies*. New Mexico: School of American Research Press, 1997.

[59] Goffman, *Frame analysis.*



are in one reality and the stage is in another. *Opening remarks* often act as bridges between these two social spaces, denoting where the frame is located.[60] Other examples of anchoring include entering a clinic: the initial moments of establishing who is the doctor and the patient, such as a lab coat. When getting pulled over by a police officer, the power interplay between the policeman and the person being pulled over is established through the badge and the action of pulling over, which anchor this activity.

When the ENCODE pilot project launched in 2003, the initial launch meeting served as the *opening remarks* of the ENCODE project. The initial framing of the project would be clarified, and anyone would be provided with the tools to see where the ENCODE frame ended and the external world began. At this meeting, many initial characteristics of the ENCODE project were set. To begin, ENCODE was framed as the National Human Genome Research Institute's next project to follow the footsteps of the Human Genome Project. It would build from understandings of the human genome, specifically that which the Human Genome Project had established. Many of the initial grants and funding opportunities from 2003 through 2004 frame ENCODE as compulsory: the obvious next step to address the "needs" left behind by the Human Genome Project,[61] or to build from the "success" of the Human Genome Project.[62]

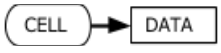
This served as an act of *frame alignment* from the very beginning of the project: instead of the Human Genome Project ending when its goals were met, it was extended further to a new project. *Frame alignment* is a term created by social movement theorists who have built from Goffman's theories. There are four established types of *frame alignment*: when frames *bridge*

---

60 Goffman, *Frame analysis:* 1974: 257.

61 National Institute of Health: RFAHG04001. https://grants.nih.gov/grants/guide/rfa-files/RFA-HG-04-001.html

62 National Institute of Health: RFAHG07029. https://grants.nih.gov/grants/guide/rfa-files/RFA-HG-07-029.html



across to other frames to link ideologies, *amplify* other frames by protecting, promoting, or clarifying another organization's goals, *extend* frames by highlighting potential ideas with which a frame is congruent, or *transforming* frames which, like *rekeying*, is when frames change shape.[63] Before the ENCODE project even launched, it could rest on the accomplishments of and support given to its predecessor. Thus, it made sense for Francis Collins, one of the Human Genome Project's leaders, to be the first to introduce the ENCODE project to the public, and when he introduced the ENCODE project, he *bridged* the two projects together, and *extended* the frame of the Human Genome Project.

Francis Collins initially stated that the ENCODE project was only interested in other organisms as they related to the interpretation of the human genome. This made the line very clear between what was accepted and what was not accepted as ENCODE activity. In fact, the purpose of selecting only 1% of the human genome to begin with was to *anchor* activity: it would create a small and unified group, and would distinguish ENCODE behavior from non-ENCODE behavior. Peter Good, a deputy director of genome sciences with the National Human Genome Research Institute, elaborated further, at this same conference, by saying that everybody working for the ENCODE project would work on the *entire* 1% region. This formed the first of the ENCODE project's anchoring frames: **only 1%.**

This 1% was called the ENCODE pilot project, and it was only the first phase of ENCODE, until ENCODE would expand and take on the entire 100% of the human genome after three years. Through ENCODE, the National Human Genome Research Institute was not only soliciting information on this 1%, but technology development that would help solicit even

[63] Snow et al. "Frame Alignment Processes, Micromobilization, and Movement Participation."

more information on this 1%. ENCODE would not accept technology development that would work for the whole genome,[64] and ENCODE policies only applied to studies done on this specific 1% section.[65] The pilot project's boundary of only looking at 1% of the human genome would allow for a clear line to be drawn between ENCODE participants and non-ENCODE participants. 1% was the first way of *anchoring* ENCODE activity. It was a clear beginning place for ENCODE to set its goals and intentions, and a ground on which to enact its policies and values.

Collins said that his understanding of ENCODE came from attempts to "take the pulse of the scientific community"[66] which had led him to ENCODE as a new frontier: the "New Edifice of Genome Research."[67] He described this "new edifice" as a building with three floors: one which relates genomes to biology, one which relates genomes to health, and another which relates genomes to society, with cross-cutting elements "like a Frank Lloyd Wright building."[68] ENCODE, as the beginning of this new edifice, was in this way set up as not only biological, but interdisciplinary. The branch which would look at the relationship between genomes and health would not look at disease, as past researchers have, but would look at health, which would allow them to establish new ideas of normative, healthy genomes. The genomes to society branch was about policy options, regarding the use of genomics in both medical and nonmedical settings. This established the responsibility of ENCODE researchers to pay attention to the reach of genomic information outside of the gene itself. Scientists could not simply hope that their

---

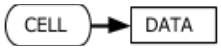
64 NIH Center for Information Technology. *Encode Project Launch:* 01:40:00

65 NIH Center for Information Technology. *Encode Project Launch*: 01:31:00

66 NIH Center for Information Technology. *Encode Project Launch*: 00:01:07

67 NIH Center for Information Technology. *Encode Project Launch:* 00:03:00

68 NIH Center for Information Technology. *Encode Project Launch:* 00:03:20



discoveries would turn out well, but they would be responsible for ensuring that their work would influence society responsibly. The rest of the launch meeting would go on to talk about methods for research, despite Collins' establishment that this was only the first floor of a three-floor Frank Lloyd Wright building.

Despite Collins' statement that ENCODE activity would be multidisciplinary, and an important project to social scientists and physical scientists alike, the room was full of genetic scientists and medical researchers.[69] The distinct lack of presence of social scientists, given Collins' statement that social scientists were integral to the project, established that the scientists here would respect and appreciate interdisciplinarity, the blending of disciplines, in concept, but that actual sociology would be a third floor of others, who would help them out, establishing more of a culture of multidisciplinarity, the acknowledgement of other disciplines. These scientists acknowledged that they were allied in a desire to be socially responsible, but that they did not actually have explicit *responsibilities*.

Collins then elaborated that matters of race and ethnicity were particularly important issues in the genome mission. This allowed ENCODE to set itself up as non-racist, but without the work or responsibility of being anti-racist by exploring those matters of race and ethnicity explicitly. He said that "at some point, we will have to deal with the questions about boundaries, and are there applications of genomics that society is telling us we should not be crossing over certain lines."[70] These lines would thus be established when ENCODE came to them. As Collins imagined it, society was external to this consortium, and would make these boundaries clear, but

---

[69] A complete list of participants was provided in the introduction chapter.

[70] NIH Center for Information Technology. *Encode Project Launch:* 00:07:40

until then, as long as ENCODE researchers proceeded with a sense of respect for its social consequences, they were doing their job responsibly enough.

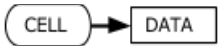
Furthermore, in his presentation, Collins established that what he was proposing in launching the ENCODE project seemed audacious, maybe even impossible. He addressed that the audience looked dumbstruck by the reach of the ENCODE project, but that the ENCODE project would not violate "any laws of physics."[71] ENCODE would continually expand as far as possible, based on its own scientific definition of possible, and society would externally regulate. Thus, Collins established that as ENCODE cast its initial anchor, it was free to roam from that point until it hit boundaries where society, an external entity to the ENCODE project, would intervene.

By only studying 1% of the genome, the ENCODE project was able to evade some of the social responsibilities of human research and establish that "third floor" to relate genomic work to society. This way it would be a founded and established project before it expanded to 100% of the human genome, and would not have to face the social consequences of studying the entire human genome until the project was better organized and sorted out. It also would serve as a more approachable goal, which would help researchers figure out if they should be ENCODE-affiliated. One panelist explained: "I would worry a lot about a group that didn't want to take on the 1% because it was too onerous, how onerous then would it be to multiply by 100."[72] 1% was ENCODE casting an anchor, but knowing that it would not stay still for long. The project would reopen membership later on, but for now it needed to close the doors and get off the dock.

---

[71] NIH Center for Information Technology. *Encode Project Launch:* 00:09:10

[72]NIH Center for Information Technology. *Encode Project Launch*: 01:54:00



One researcher, Eric Green, who had recently become the director of the National Human Genome Research Institute, was called up to present on the work of the lab to which he was the primary investigator. This was not planned, but the panelists found that having an example of a researcher present would probably be helpful for the sake of "laying the groundwork" for the ENCODE project.[73] His work primarily used back libraries, a method practiced by ENCODE researchers using sequences from animals in which long stretches of DNA were completely homologous[74] between the animal and the human. He estimated that 30 megabases (1% of the human genome) would be a possible goal for the sequencing of three species for back libraries a year. One audience member addressed Eric Green directly to argue that ENCODE's goal was impossible. He asked Green if he could do what his lab was doing now several times faster, to meet the goal of the ENCODE pilot project. Green replied that he certainly could not, but that by the end of the three year pilot project, he was confident that this would change.[75]

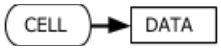
ENCODE in this way was framed as a biological project that could change what was possible in the realms of biology and technoscience, but could not change what was possible in terms of certain issues deemed 'social,' including race and ethnicity. This formed the second anchoring frame of the ENCODE project: it was **developing new possibilities** for genome research. It would proceed to change the world of biology, but would proceed with caution, waiting for social entities to stop it or call it out.

---

[73] NIH Center for Information Technology. *Encode Project Launch*: 00:57:35

[74] *Homologous* is a term for when two chromosomes (which are essentially bundles of DNA) have the same structure and features. It is often used in genetics when chromosomes are aligned during reproduction or, in this case, when segments of DNA from two or more species lines up.

[75] NIH Center for Information Technology. *Encode Project Launch*: 01:24:00



Furthermore, ENCODE was a project concerned with the human and only the human, forming the third anchoring frame: ENCODE was **humans only**. It would not use other organisms, or support the development of technology to be used on other organisms. One audience member expressed confusion with this policy, and said that the only way to look at gene expression, and establish genes as functional, would be to look at their expression when a gene is either bred out, or turned on or off, which is not possible to do in studies of humans, because humans cannot be bred.[76] He was met with two responses. One was that this was the point of the ENCODE project: ENCODE was going to do the impossible, by building technologies in a domain where technology was limited. In this way, ENCODE was framed as a project that was not limited by preexisting ideas of what is possible in the field of genomics research. It asked participants to think beyond what was possible now, and imagine a world where new things were possible. Some research would have to be different in humans than in model species, because humans are fundamentally different from other organisms.

The other response was that this boundary between the human and the animal was admittedly quite unclear, and would remain flexible:

> "It is hard to draw this bright line that we are asking people not to cross in terms of organism functional studies. But if you are going to propose something of that sort it needs to scale extremely well, it needs to be quite financially affordable, and it needs to be clearly done in a way that informs the human sequence, which is the focus of this enterprise." (NIH Center for Information Technology, 1:22:52)

Thus, ENCODE's anchoring was firmly set in the *human* genome, even if the parameters where human stopped and animal began, remained unclear and flexible. This was an integral part of

---

[76] NIH Center for Information Technology. *Encode Project Launch:* 01:21:24

how the ENCODE project would accomplish its task of creating new possibilities in genomic science.

In staying linked to the human, ENCODE was able to maintain its *frame alignment* with the recently finished Human Genome Project. While the ENCODE project may have picked up some of the controversies that surrounded the Human Genome Project, it was able to evade any controversy about the use of model species. It was also able to define itself as entering uncharted territory: by forcing the audience member who could not conceptualize methods that ethically use human subjects, and no animal subjects, to rethink his methodology, ENCODE was setting itself up as a project which would not only *use* what was possible to understand DNA, but *develop* new possibilities of how to understand DNA.

One exception to the humans only rule was the use of animal 'back libraries:' the comparison of homologous stretches of genetic code from more than one species. That stretch of DNA could be studied in the animal instead of the human, only if to confirm a study done in a human, or if the only methods proposed to study that code were not possible in humans. The use of back libraries entails the splicing of pieces of animal's codes together to form a frankenstein-esque composite organism resembling a human. Each piece can be traced back to another organism, but they are stripped of context and refitted to form the pieces of a human. Feingold explained that it is common that the National Human Genome Research Institute will receive requests, called *white papers*, for a specific animal to have its genome sequenced because this sequence will be helpful for the study of human genomes.

A second audience member, named Bruce, from the University of Oklahoma, expressed concerns that the only ways he knew how to identify gene expressions in humans throughout the

whole body relied on the use of stem cell lines, and "the president[77] has his own agenda on dealing with that issue."[78] This indicated that research on humans was too heavily politicized to use the technology that was in his vernacular, as public funding of stem cell use was already banned at the time. He then went on to say that ENCODE needed stronger boundaries, and a clearer set of limited stages before growing to the point that it was creating an entire comprehensive encyclopedia of the whole human genome. As he finished his critique, he joked that he himself was being too political: "I must excuse myself or I'll never get funding again."[79] The audience and panelists laughed. Even though the project was defined as interested in investing in the impacts of genomics to society, Bruce's fear of being political indicated a precedent that scientists should not be political- politics should be handled instead by the metaphorical 'third floor.' The response to Bruce's comment was that he was thinking too big: "the basic idea of this effort is to define the parts. We are not trying to figure out functions of all the parts."[80] His political edge was excused because the project was going to start smaller, meaning that it was yet to hit those political boundaries.

In order to accomplish this goal of changing what was possible in the world of genetics research, the ENCODE project would blend both computational and 'wet-bench,' or experimental, procedures. Computational procedures typically refer to gene prediction methods, in which enough information about how genomes work is used to assess the most likely processes through which genetic codes are transcribed. Most methods of computational gene

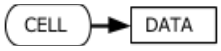
---

[77] George W. Bush at the time.

[78] NIH Center for Information Technology. *Encode Project Launch*: 01:57:00

[79] NIH Center for Information Technology. *Encode Project Launch*: 01:58:33

[80] NIH Center for Information Technology. *Encode Project Launch:* 1:58:46



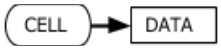
prediction rely on understanding a an alphabet-like quality of DNA: if you can understand the words most likely to be written in genetic code, and the process through which letters become words, you can take a string of DNA which has one or a few single point mutations (when one letter has been switched out for another) and interpret what the word was going for, and thus understand evolutionary elements.[81] Experimental procedures, on the other hand, are far more common and can refer to a very diverse array of methods. They may refer to any method which uses actual laboratory space and observations/manipulations of physical materials. The act of blending these two procedures thus questions a boundary that had previously informed a lot of genetic research and significantly changes the way a molecule leaves a cell and becomes data. This is the fourth anchoring frame of the ENCODE project: ENCODE combines **computational +experimental.**

This combining of experimental and computational procedures caused confusion at the 2003 launch meeting. Some audience members were confused as to how the same standards of reliability can be applied to both. If two techniques with two different standards of reliability are under the same frame, the frame would be devalued. One panelist met this confusions by envisioning a circular relationship between the two: data would be provided through experimental procedures and then computational procedures would look at the same regions and see if they validate the findings of the experimental procedures, or vice versa. He explained "this is the closed loop of the experimental work feeding back on the computational work, the computational work providing predictions to the experimental work."[82] In fact, the National

---

[81] Michael R. Brent. "How does eukaryotic gene prediction work?" *Nature biotechnology* 25, no. 8 (2007): 883.

[82] NIH Center for Information Technology. *Encode Project Launch:* 01:35:09



Human Genome Research webpage for ENCODE describes this conversation between computational and experimental researchers as one of the primary goals of forming the ENCODE consortium.[83] ENCODE would be the product of the conversation between the two; they would not only be aligned, but they would mix in a reflexive feedback loop.

Finally, throughout many of these anchoring statements, the ENCODE project was described as limited by a decisive plan to only look at *functional* DNA elements.[84] This was one of the first times ENCODE users began to use their own definition of the word *functional*, which will come up in chapter four as a large controversies surrounding the ENCODE project, but was even stirring confusion as early as 2003. The official Requests for Applications, documents to state official grant opportunities, announced at the launch meeting were called "Determination of All Functional Elements in Human DNA"[85] and "Technologies to Find Functional Elements in Genomic DNA."[86] Elise Feingold, ENCODE founder who introduced the Requests for Application said that "there may be a need to do a small analysis of the function… but any in-depth study of functional elements… is really beyond the scope of this RFA."[87] Function was established in this moment as a crucial word in translating DNA to the world, and formed the fifth and final ENCODE anchoring frame: ENCODE was only concerned with **functional** elements. It would not only set limits on which DNA segments were studied by ENCODE, and thus which DNA segments mattered to ENCODE, but it was a foundational assumption of what

---

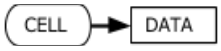
[83] National Human Genome Research Institute. "ENCODE Project Background." *NIH National Human Genome Research Institute.* (Last Updated: May 21, 2012). Accessed: March 2017.

[84] This might seem contrary to the ENCODE frame of looking at the entire human genome, but the idea was that the entire human genome would be seen as *candidate* functional elements, and then individual sequences would be analyzed and validated based on whether or not they had function.

[85] National Institute of Health: RFAHG03003 https://grants.nih.gov/grants/guide/rfa-files/RFA-HG-03-003.html

[86] National Institute of Health: RFAHG03004 https://grants.nih.gov/grants/guide/rfa-files/RFA-HG-03-004.html

[87] NIH Center for Information Technology. *Encode Project Launch:* 00:26:26



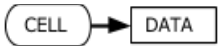
DNA was and what DNA does. No specific functions of DNA were discussed at this point, it simply set up a theoretical dualism: functional DNA does something, while nonfunctional DNA does nothing. The Human Genome Project was a project which established that DNA exists, and a two-dimensional idea of what it looks like. ENCODE, on the other hand, was founded on an established idea of what DNA *does*.

The boundary of only looking at functional genes was loose and largely undefined. One audience member asked if a part might not necessarily be functional, but could be useful for research purposes in another way, should all the parts still be annotated as part of the ENCODE project? Two panelists replied "all the parts:"[88] one expressed this with uncertainty, and one expressed this jokingly, but neither speaker elaborated. The boundary of functional DNA versus non-functional DNA was beginning to be poked at already, giving these presenters the option of either defining function clearly and explicitly, or letting go of it. They let go of it somewhat, but continued to use it to anchor the ENCODE project.

In summation, the 2003 launch meeting established five original anchors for the project: it would be an agent in the changing possibilities of genomic biological research, with limits imposed by other disciplines (developing new possibilities); it would only look at humans, or that which was compellingly applicable to humans (humans only); it would only look at 1% of the genome, for now (1%); it would combine computational and experimental procedures, and the standards that came along with both (computational+experimental); and it would only look at "functional" DNA elements, even though function remained largely undefined (function). These limits referred to who would receive ENCODE funding and be accredited on the National

---

[88] NIH Center for Information Technology. *Encode Project Launch:* 02:06:31



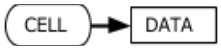
Human Genome Research Institute, as well as who was a member of the consortium, which would enforce open channels of communication between participants. An audience member commented on people who wished to participate but acquired funding from other sources and maybe did not abide entirely by these restrictions, to which Feingold replied that ENCODE data would be available for external use, but if people are not willing to abide by these criteria, they are not a part of ENCODE. Another audience member asked if letters of intent could be enough to establish collaboration with the ENCODE project, to which a panelist responded that no, ENCODE "can't match-make."[89] The meaning of ENCODE participation was, in this way, firm in that it could not be taken lightly, but vague in that the each of these anchors had its own flexible characteristics.

ENCODE Expands

In 2007, ENCODE had its first major expansion. Phase one, which looked at only 1% of the genome, was complete. EGASP, a project developed for the assessment of the ENCODE pilot project, was used to assess that accuracy of the findings produced through the ENCODE pilot project, and officially presented its results in 2006. Thus, ENCODE took EGASP's feedback and launched phase 2: the complete ENCODE project to study 100% of the human genome. The grants released by the National Human Genome Research Institute at this time no longer needed to be aligned with the Human Genome Project because the ENCODE project was established, as both a consortium and a research project. Although it still rested on the information that the Human Genome Project had produced, it was time for ENCODE to have its own identity and set its own frames.

---

[89] NIH Center for Information Technology. *Encode Project Launch*: 02:19:50



To accommodate this expansion, the National Human Genome Research Institute released a request for applications to form a Data Coordination Center, to keep ENCODE data organized, and Data Analysis Center, to facilitate the "integrative analysis"[90] of data produced by the ENCODE project.[91] The Data Coordination Center became hosted at Stanford through Michael Cherry's research group, and the Data Analysis Center became hosted at University of Massachusetts Medical School, Worcester through Zhiping Weng's research group. The Data Coordination center would go on to facilitate many large institutional changes in the ENCODE project, including forming new grants and subcommittees, and help facilitate some of the ENCODE project's engagements with non-ENCODE researchers. The Data Analysis Center operates on a much less public basis, but translates ENCODE data into the 'encyclopedia:' an interactive online matrix that sorts all of the findings of the ENCODE project.

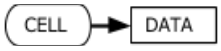
The GENCODE Consortium was also launched around this time, which supplements ENCODE by providing tools for analysis and annotation of ENCODE gene features. It began by annotating and validating the results of the pilot project and following ENCODE's expansion. It integrated the work of UK-based research groups HAVANA (Human And Vertebrate ANalysis and Annotation), which provided manual gene annotation, and *Ensembl*, which provided computational gene annotation. One of GENCODE's main goals was to merge the manual and the automated, to provide a tool that would be helpful for the comprehensive and exhaustive breadth of the ENCODE project.[92] This was part of the expansion of ENCODE's goal to

---

[90] National Human Genome Research Institute. "ENCODE Participants and Projects."

[91] National Institute of Health RFAHG07010 http://grants1.nih.gov/grants/guide/rfa-files/RFA-HG-07-010.html

[92] Jennifer Harrow, et al. "GENCODE: the reference human genome annotation for The ENCODE Project." *Genome research* 22, no. 9 (2012): 1760-1774. [Chapter four will discuss controversy formed around whether or not this is actually an accurate way to describe pseudogenes]

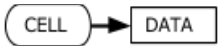


combine experimental and computational procedures. Through the emergence of GENCODE, ENCODE was able to further build upon and expand its practices of merging the computational and the experimental. Where the pilot project had put the two on the same plate, GENCODE put them in a blender, and they could no longer be teased apart.

In order to blend together the experimental and the computational, GENCODE had to develop new "controlled vocabulary,"[93] and establish new sets of validating procedures which would be reassessed every three months. GENCODE not only established its own language, but set its own standards of what makes genetic studies reliable by establishing a practice of *inter-institutional consensus making*: rather than having an individual lab or research group publish findings based on their own standards of replicability or a publisher's standards of replicability, to prove data valid enough for release in ENCODE databases, researchers could use the tools provided by GENCODE, a larger and more expansive consortium representing and supporting multiple institutions. These tools put many different types of research at different institutions in conversation with one another, making them not only hybrids of computational and manual procedures, but hybrids of procedures from a variety of institutions. Recall that when it was first proposed that ENCODE would blend experimental and computational procedures, conflict arose over the idea of combining these two systems of validity. This conflict was met with the response that they would not simply be combined, but put into a cyclical conversation. The GENCODE consortium held this conversation, not only by putting these two types of research methods into conversation, but by opening up many institutions to participate in that conversation, and letting the differences between the two become indistinguishable.

---

[93] Harrow et al. "GENCODE: the reference human genome annotation for the ENCODE Project."



Furthermore, one of the GENCODE consortium's main tasks was to decipher which genes were functional versus which genes were 'pseudogenes,' seemingly functional genes that had 'lost' their function due to evolution.[94] GENCODE validated findings of genetic researchers by affirming that genes had functions through inter-institutional consensus making. GENCODE was not only validating the findings of individual labs, but it was validating ENCODE's overall practice of setting the boundary of 'functional' genes. Thus, the act of blending between the computational and experimental, which informed the creation of the GENCODE consortium, emerged as a method through which the previously undefined concept of 'function' took shape. As the line between manual and automated was blurred, a new line between functional and nonfunctional was increasingly strengthened.

Finally, two projects to specifically look at model species emerged at the time of the 2007 ENCODE expansion: modENCODE and modERN (ERN for Encyclopedia of Regulatory Networks, which refers to the regulatory elements that shape how DNA is transcribed and forms proteins). These two projects would not only facilitate the move from 1% to 100% of the human genome, but they would provide model species resources to ENCODE's analyses. Much like GENCODE, they were separate institutions and separate consortiums, but which existed for the sake of the ENCODE project. They would use the models of *Drosophila* (fruit flies), and *Caenorhabditis* (nematodes).

Although these consortiums are referred to as projects for the study of model species, the ENCODE consortium with its human-centric mission came first. Human-centrism was the

[94] Harrow et al. "GENCODE: the reference human genome annotation for the ENCODE Project."

master frame whose *scope* extended beyond the human.[95] Thus, ENCODE uses the phrase *model species* in an entirely unique way, that challenges the relationship between humans and models in a way much like the flippant comment at the 2003 launch meeting: "unless you want to call *us* a model organism."[96] The framing of animals as models through the designation of animals as modENCODE and humans as the main ENCODE, shows that the ENCODE project does not actually take humans as the models for animals in terms of the project's larger scope, but it did allow a human-based project to model for an animal-based project, and human-based technology development to model for animal-based technology development.

The line between human and nonhuman took a new shape. Humans did not build from models, and models did not build from humans, but models and humans were reciprocal and supplementary to one another. Much like how the experimental and computational procedures were initially put in conversation to create ENCODE as the product, human and nonhuman genomes were put in conversation to create a broader ENCODE project. The complete form of ENCODE's genome resembled a human and related to humans, but pieces of it did not come to be as humans originally, or follow the evolutionary pathways from which human life emerged. It self-identified as human, but with flexibility. It could achieve that which a human cannot achieve, in that it could also achieve what *Caenorhabditis* and *Drosophila* do. As it grew, it would pick up new animalistic capacities on the way, as it will with the 2011 emergence of mice in the ENCODE project.[97]

---

[95] David A. Snow, E. Burke Rochford Jr., Steven K. Worden and Robert D. Benford. "Frame Alignment Processes, Micromobilization, and Movement Participation" American Sociological Review, Vol. 51, No. 4, August 1986 p 464-481.

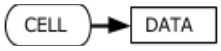
[96] Snow et al. "Frame Alignment Processes:" 464-481.

[97] Feng Yue, Yong Cheng, Alessandra Breschi, Jeff Vierstra, Weisheng Wu, Tyrone Ryba, Richard Sandstrom et al. "A comparative encyclopedia of DNA elements in the mouse genome." *Nature* 515, no. 7527 (2014): 355-364.

In expanding from 1% of the human genome to 100% of the human genome, ENCODE *rekeyed* for what it meant to create an encyclopedia of human DNA elements. Procedures for experimental and computational analysis were blended by the GENCODE consortium, and studies of humans and model species were blended by the modENCODE and modERN consortium's emergence in an act of *upkeying* that invited animals into the frame of the human genome. The information which ENCODE presented thereafter was framed by these relationships as a hybrid for which the lines of human/animal and organism/machine that apply to anything outside of ENCODE are irrelevant. These words now had their own-ENCODE specific connotations, and ENCODE was not only exploring the human, but defining it.

To expand upon Goffman's metaphor of the anchoring of activity, when ENCODE was only looking at the *functional* genomic elements of *1%* of the *human* genome through combination *experimental* and *computational* procedures, it was a boat with a much smaller rope connecting it to its anchor, firmly set in these five qualities. When it re-launched as a project which looked at 100% of the human genome, its rope was 100 times larger, making it harder to see where the anchor exactly ended up: somewhere where the previous dualities of human/ animal versus organism/machine took on a new meaning. Its new, longer rope provided new mobility and new limits to how far it could travel, which led to a *rekeying* of ENCODE activity.

Upkeying occurred when ENCODE expanded to 100% of the human genome, and model species entered the ENCODE project, making ENCODE's framing of *humans only* expand to include *Drosophila* and *Caenorhabditis.* The framing of *experimental+computational* procedures remained, but got stronger and more solidified through the GENCODE consortium which not only combined the experimental and computational, but created a new system which deeply



blended the two into a new form of consensus-making. The keyings of *functional* genomic elements and of *developing new possibilities* will continue to change form and shape throughout the following years of the ENCODE project.

Seeing DNA Through these Frames

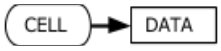
Donna Haraway explored the way the human genome project and the mouse genome project (hosted by the Sanger Institute in the UK) pointed towards a new future with new designations of what was human and what was nonhuman. On a material level, the ENCODE project moved the genome out of the cell and into the form of cybernetic data, and the social arrangements in which it takes shape are that of a cybernetic society, though it has implications for a corporeal society as well. Prior notions of what shapes society in a material world cannot directly apply to a cyber world.[98] Thus, approaching ENCODE beginning with a frame analysis has allowed me to start with nothing and build the world of ENCODE from scratch. Now that I have shown the framing of ENCODE, let us revisit Haraway's question from the beginning of this chapter: what has enabled ENCODE's understanding of DNA?

Caenorhabditis and *Drosophila* serve as model species for humans because a significant portion of their genome is understood to match with the human genome.[99] This is the perspective modENCODE and modERN take, and that ENCODE has taken since the introduction of back libraries in 2003: much of the human genome is the same as the genome of so many other

---

[98]Donna Haraway, "Mice into Wormholes."

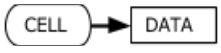
[99] If you are unfamiliar with this concept, you might be thinking: 'but humans do not share half of our traits with fruit flies.' This is explained by epigenetics, the idea that genes are turned 'on' or 'off,' or by the idea that many of those genes we share are *non-functional*, both topics which ENCODE continues to explore and which will come up again in the following chapters. Essentially, the same nuanced concepts of genetics and gene regulation that the ENCODE project is focused on are these concepts which feed the idea that two species can have 50% or more matching genomes without having close to 50% of the same traits are the concepts that the ENCODE project focuses on.



animals. These animals are not quite model species for humans, they *are* humans, if you only look at them in pieces. And how could you not look at them in pieces? The human genome is roughly 3.3 billion letters long. According to ENCODE director Mike Pazin, that is a thousand copies of *War and Peace.*[100] It is hard to find technology that can store all of that data in one place, so for genomics researchers, the human genome is more often than not, fragmented. It might be intuitive that looking at the genome in pieces (files, datasets) makes it easier to see than looking at it as a comprehensive genome, but recall how figure 1.1 in the previous chapter showed that the ENCODE project is *building* from the human genome. It is taking the two-dimensional string of letters that form the human genome and making them three-dimensional, making it countless times harder to look at because of its newly developed width and height.

What ENCODE is seeing that allows the relationship between human and fruit fly to be reciprocal, rather than one modeling for another, is that there is no human, there is no model, there are only pieces. The entire concept of species, or debates over evolution, do not matter in any of the procedures, experimental or computational, run by ENCODE, because they are only looking at a small stretch of code, code which is simultaneously life (the subject of experimental study) and technology (the subject of computational study), and which is fragmented to the point that there is no physical, let alone social, form visible. These codes are not embodied, they do not even resemble traits, making species an irrelevant category. Thus, in 2007, when the GENCODE consortium blended together the manual and computational, it was precedented because the genome was already so distanced from physical-form DNA. In 2013, when mouse DNA entered

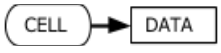
[100]Mike Pazin, "Can an 'Orchestra' of Scientists Find the Hidden Music in your DNA?" from TEDMED, October 28, 2013. http://www.tedmed.com/talks/show?id=103852



the ENCODE project, it was precedented because of this reciprocal relationship between humans and other animals that allowed the *humans only* frame to be rekeyed.

ENCODE is looking at DNA which is so large that everything, including the ENCODE project itself, is tiny by comparison. To be read, it must be broken down into very small pieces, meaning that the only genome anyone can see is very small pieces. The ENCODE project spans the entire DNA molecule, which is at once a thousand copies of *War and Peace* and at once microscopic. These segments are so minuscule in the context of the broader genome that no individual could possibly see how they form together to create an organism, whether that organism be a worm or a human. Each individual human researcher contributes a minute, microscopic piece to a much broader project. Exploring 1% of the human genome may seem like a small endeavor compared to the scale-up project, but within that 1% lay many, tiny pieces.

Earlier in this chapter, I discussed Francis Collins' idea that the ENCODE project would interact with society as an external regulator, and that race and ethnicity were crucial matters for the ENCODE project to discuss. If the genome is fragmented into pieces too small to distinguish between a mouse and a human, how could it speak to race or ethnicity, the differences between humans? As of 2007, the genome which ENCODE looked at had yet to find its place in a world of people, where they interact with others in social capacities, because the genome was just moving into cyberspace, away from physical bodies. Eventually, however, ENCODE would produce not only data, but findings, and findings, as the three-story model discussed earlier in this chapter indicates, will eventually enter the worlds of biology, medicine, and society.



CHAPTER 3. THE MOLDABLE GENOME.

Introduction

In the previous chapter, I introduced some vocabulary for frame analysis, and introduced my approach of hybridizing Actor-Network Theory and frame analysis. I used these tools to discuss the 2003 ENCODE launch meeting as the anchoring of ENCODE activity. It ultimately showed the introduction of five qualities with which the ENCODE launch project was framed: *only 1%*, *developing new possibilities*, *humans only*, *computational+experimental*, and *functional*. In the first major expansion of the ENCODE project, the 2007 scale-up project,[101] ENCODE destroyed the framework of only looking at 1% of the human genome, and replaced it with looking at 100% of the human genome. It further blended together experimental and computational research methods through the GENCODE consortium, which used *inter-institutional consensus making* to form information beyond a binary of computation and experimentation. It also began to rekey the frame of *humans only* by introducing the modENCODE and modERN consortiums.

In this chapter, I will discuss the ways ENCODE has maintained the frame of being an agent in changing possibilities, and ENCODE's overarching master frame,[102] which I have not yet addressed, of being an *Encyclopedia,* as the project entered phase three in 2012. However, instead of following DNA through ENCODE's frames, I will follow data, DNA's next iteration in the ENCODE continuum from cell to data to encyclopedia and beyond.

---

[101] Also known as ENCODE phase 2 or ENCODE 2.

[102] Snow et al "Frame Alignment Processes."

Part of the 2007 scale-up project was the establishment of a Data Coordination Center, based at Stanford University in Stanford CA, and a Data Analysis Center, based at the University of Massachusetts Medical School in Worcester, MA. Members of the Data Coordination Center are referred to as curators or data wranglers on the encodeproject.org website.[103,104] These institutions would allow ENCODE to hold and manage the larger portion of data that the ENCODE project would now produce, by regulating and presenting data. In 2012, the Data Coordination Center established a website for the ENCODE project, encodeproject.org. Until 2012, all data was hosted on websites, called genome browsers, that were not ENCODE-specific, and held data from a variety of genomics projects. Encodeproject.org became an entirely new channel through which ENCODE data would be communicated to users. The encodeproject.org website displays data completely publicly. All that is required in order to download any portion of the ENCODE project's data is to be in a country with access to this website, which currently includes all of North America, Europe, and Asia. Yet, there are specific populations which the ENCODE project frames the website to be *for,* by referring to a 'scientific community' and a body of 'users.'[105]

In the previous chapter, I mentioned that an Actor-Network Theory analysis may provide insight towards how ENCODE translates DNA to the larger world. This elucidates the middle spaces between the insights of the sociology of The New Genetics, which frequently focused on a nature/nurture dichotomy between DNA and sociological perspectives, and the DNA molecule

---

103 "Acknowledgements" *Encode: the Encyclopedia of DNA Elements*. Accessed November 2016. https://www.encodeproject.org/acknowledgements/

104 These roles were described in Chapter One.

105Software is developed for the 'community' released based on its benefit to the community 'community' according to the encodeproject.org website guidelines:
"Data Use Policy" *Encode: the Encyclopedia of DNA Elements.* Accessed November 2016. https://www.encodeproject.org/about/data-use-policy/

itself, which is far more complex and nuanced. The first step along that process was to look at how physical-form DNA is translated into data, which the last chapter explored through the ENCODE project's initial launch and methodologies, such as the use of model species (or lack thereof), or the use of computation or experimentation. In this chapter, I use an Actor-Network Approach to see how that data is translated to a world of researchers and geneticists writing analyses of ENCODE findings. These are the processes through which raw data is displayed, curated, and analyzed through the ENCODE Data Coordination Center and Data Analysis Center.

Thus, to approach the question of how data analysis is framed, I ask how the ENCODE project redefines words like *encyclopedia, analysis,* and *users* through data presentation and data release practices. In order to do this, I must let go of *a priori* notions of what these words mean, and allow them to be built fresh as the website, encodeproject.org, and the actors it links me to, defined them. I then may trace what happens when data produced within the ENCODE project's frames are released, and what release even means in the first place. Ultimately, I have come to see that the ENCODE encyclopedia is a tool for visualizing DNA as a moldable and customizable, publicly available resource. While my research took place entirely online in the digital matrixes of ENCODE, if the ENCODE project is an agent in changing what is possible in the world, its online life is likely representative of its offline implications.

Along the way, I find that in order to create an encyclopedia, the body is, in many ways, fragmented into pieces, which provide new ways of looking at DNA, and subsequently at the bodies which DNA does, or can, form. With a rise in gene editing technology, the presentation of DNA as a re-shapable substrate may indicate that the ability to curate a genome has implications

for both digital and organic worlds. Thus, an Actor-Network Theory approach which blends the material and the semiotic, the physical and the ideological, can point towards what ENCODE's users see when they look at ENCODE data.

The Beginning of Data Release

The story of how encodeproject.org defines data analysis begins in 1991, 12 years before the project's launch, when the National Human Genome Research Institute and the Department of Energy together established a data release policy that called for all data and materials related to the Human Genome Project to be released within 6 months of their generation. By enforcing data release as a separate process from publication, this policy put into practice a distinction between the data that is published and the data that is released: published data has been interpreted, analyzed, and put in context. Published data typically will reference its applications and uses, using the classic structure of scientific papers that includes an introduction, methods, results, and discussion. Released data, on the other hand, has relinquished all say over how it is used, and lacks the introduction or discussion elements that contextualize it. Researchers who have generated data and released it have no say over how it is used or applied. Responsibility has been removed from their hands, and knowledge that they once held publicly has been put in the hands of ENCODE users.

When this initial policy was in place, researchers could claim ownership over their data for a six month period, allowing them to publish analyses of their data before anyone else, even though all data would eventually be released to the public.[106] They could quickly have (or at least

---

[106] National Human Genome Research Institute. *Reaffirmation and Extension of NHGRI Rapid Data Release Policies: Large-scale sequencing and Other Community Resource Projects.* National Human Genome Research Institute. February 2003. https://www.genome.gov/10506537/

submit for) a paper published under their name, in which they hint towards the applications of their paper, or interpret it for what it means, or could mean, before anybody else. They could also take this time to prepare for a finding to be significant in a way they did not want to be associated with, such as if a finding challenged a previously held belief. These six months were a period of time for researchers to be concerned with what to do with *their* data, as researchers still had ownership over their work. By ownership, I mean control, rather than authorship, which denotes accreditation. While researchers still were accredited for their work for the most part, after six months, the data was out of their control, and anyone could publish papers using this publicly available data.

This policy was called *rapid release.* Rapid release was not only a policy but a value: it said data should not be held onto for too long, and that data should be produced in a way that is considerate of the community by ensuring that it was easy for researchers to know what the others are doing. For the Human Genome Project, it was a way to ensure that all information was shared, even if it would never be published in its entirety.[107] The politics of publishing would not have an effect on the Human Genome Project's overall sequence: no longer did it matter if data was interesting enough to be published, or written about well, or who it was written about. Instead, it just had to be produced quickly. Within the guidelines for this 1991 policy it was stated that "it is also desirable to make the fruits of genome research available to the scientific community as a whole as soon as possible to expedite research in other areas."[108] Researchers

---

[107] National Human Genome Research Institute. *NIH-DOE Guidelines for Access to Mapping and Sequencing Data and Material Resources.* National Human Genome Research Institute. Last reviewed March 9, 2012. https://www.genome.gov/10000925/access-to-mapping-and-sequencing-resources/

[108] National Human Genome Research Institute. *NIH-DOE Guidelines for Access…*

could stop data production that they felt was redundant, or expedite data production that would build from their peers'.

Rapid release at this time, however, would not be *too* rapid. The advisors of the National Institute of Health and Department of Energy kept the six month moratorium[109] to give scientists a chance to verify their data's accuracy and ensure their rights to intellectual property protection.[110] Researchers could still apply for patents, or publish papers, related to their data before it was released to the public.

In 1996, during the middle of Human Genome Project, the International Human Genome Sequencing Consortium adopted a set of principles known as the Bermuda Principles. These "called for the automatic, rapid release of sequence assemblies[111] of 1-2 kb[112] or greater to the public domain"[113] within 24 hours of their generation. The Bermuda Principles were updated in a meeting sponsored by the Wellcome Trust institute in 2002 to accommodate for the increased scale of DNA sequencing[114], and in 2003, when ENCODE launched.

From 1991-2003, the Bermuda principles were established as a value, which says that knowledge is for the public, that data would be produced quickly, but without the motive of competition. Data production would not be fast because of competition or the preservation of intellectual properties, it would instead be produced quickly because of a principle, a shared

---

[109] A moratorium is a delay or suspension of activity until further notice, or until a specific goal is accomplished.

[110] National Human Genome Research Institute. *NIH-DOE Guidelines for Access…*

[111] Sequence assemblies are a specific type of data in which portions of the DNA sequence are illuminated.

[112] Kb is a unit of measuring genomic sequence. Short for kilobases, it refers to a thousand 'letter's of genomic code (A's, C's, T's, or G's).

[113] National Human Genome Research Institute. *Reaffirmation and Extension of NHGRI…*

[114] NHGRI Rapid Data Release Policy. National Human Genome Research Institute. National Human Genome Research Institute. February 2003. https://www.genome.gov/page.cfm?pageID=10506376

value to the community surrounding the Human Genome Project, and soon, the ENCODE project.

The Bermuda Principles Enter ENCODE

Actor-Network Theory calls for a material-semiotic approach, which follows actors, human or otherwise, through networks. In this situation, I take the Bermuda Principles, or the general principles of rapid release, as actors which move to enter the ENCODE frame, a crucial element to my following of ENCODE data through ENCODE frames where they meet DNA, another actor, and together release data through the framework provided by encodeproject.org. To follow these principles through the ENCODE project, I begin again at ENCODE's launch meeting in 2003.

At this meeting, one speaker presented on ENCODE data release, and another presented on data management in the ENCODE project. Mark Guyer, a leader of the Human Genome Project who took on much responsibility in the forming and founding of the ENCODE project, presented on data release. He said that "the project aims to function openly, making all data available to the scientific community in a timely manner. A timely manner is not meant to obscure anything, but is meant to be proactive."[115] He went on to say that the consortium would likely eventually establish its own data release policy for which all participants would abide. He said that while the Bermuda Principles still applied to ENCODE data, it was possible that with the unpredictability of ENCODE, different types of data would challenge these principles.[116]

---

[115] NIH Center for Information Technology. *Encode Project Launch*: 00:46:25

[116] NIH Center for Information Technology. *Encode Project Launch:* 00:48:00

An audience member took issue with Guyer's ambiguity in introducing these principles. He said that these principles need to be "uniform and really really clear,"[117] rather than to just decide what the expectations are for proper scientific behavior. Guyer replied that he agreed "with the *desire* to be as crystal clear as possible,"[118] though he was not sure it if would be possible to come up with universally applicable language. Another speaker said that this policy was "a moving target" and that there would need to be continual discussion, collaboration, and assessment.

Peter Good, one of the directors of the National Human Genome Research Institute, presented on data management. He said that a subcommittee had formed to address matters of data management, and that there were going to be issues with data management that could not even be predicted or foreseen.[119,120] To tackle these challenges, they would try to leverage a variety of existing databases, rather than immediately develop a new type of database, or 'genome browser,' a particular type of database specific to genomics projects, that allows a user to scroll through the entire sequence of DNA.

By leveraging existing technologies, the ENCODE pilot project could focus on data production, and leave data management to be dealt with later on. The presenter on data management said that too many different types of data would make data presentation "difficult and cumbersome,"[121] meaning that both data management and presentation would be worried

---

[117] NIH Center for Information Technology. *Encode Project Launch*: 00:49:46

[118] NIH Center for Information Technology. *Encode Project Launch*: 00:50:32

[119] NIH Center for Information Technology. *Encode Project Launch:* 01:06:12

[120] The data management challenges would include: displaying diverse types of data, developing standards for such data, keeping data linked to the sequence produced by the Human Genome Project, and making data easily accessible.

[121] NIH Center for Information Technology. *Encode Project Launch:* 01:07:25

about at a later date. For now, the speaker explained, it was most important that data was easily accessible, rather than organized or edited. The subcommittee for data management decided to begin with using the University of California Santa Cruz genome browser for sequence-based data, and a database for 'other' types of data. Soon it added on the the Gene Expression Omnibus,[122] run by the National Center for Biotechnology Information, and the ArrayExpress archive of functional genomics data, run by the European Bioinformatics Institute, for 'other'[123] types of data.[124]

In 2004, all sequence-based data and findings became stored on the University of California Santa Cruz genome browser, a web portal that has been publicly accessible since 2001. This is a portal that held genomic information for a variety of browsers, including humans and many other animals, as well as yeasts and the ebola virus.[125] A user can choose which animal they wish to focus on, from which they may compare and contrast other genomes. A user may search for a human genome and click on a specific part, and find that they have been directed to the genome of a non-human. All animals available for browsing are shown on the homepage in a phylogenetic tree that allows their evolutionary relationships to be seen.

The University of California Santa Cruz genome browser, which holds the majority of ENCODE data, enforced a policy specific to ENCODE that included a nine-month moratorium that allowed release to be distanced from publication. The purpose of this policy, which was

---

[122] NCBI. "Gene Expression Omnibus." *National Center for Biotechnology Information.* Accessed March 2017. https://www.ncbi.nlm.nih.gov/geo/

[123] 'Other' data referred to that which did not come directly from the DNA sequence, but that looked at other genomic elements, such as data on gene expression, which are a major part of the ENCODE project's research.

[124] EMBL-EBI. "Array Express- functional genomics data." *ArrayExpress*. Accessed March 2017. http://www.ebi.ac.uk/arrayexpress/

125 University of California Santa Cruz. "Genome Browser Gateway." *UCSC Genome Browser.* Accessed March 2017. https://genome.ucsc.edu/cgi-bin/hgGateway

released as a guideline on the University of California Santa Cruz genome browser, was to allow data producers the opportunity to publish their findings before releasing them, if they wished.[126] In this way, the ENCODE project maintained a strict division between data release and publication, and the Bermuda Principles remained upheld for ENCODE data, even in this new genome browser framework. Without its own autonomous genome browser, the Bermuda principles of rapid release were negotiated with the University of California Santa Cruz or other genome browsers. Thus, ENCODE data was subjected to moratoriums, compromises between immediate rapid release and no rapid release.

Phase Three

In 2012, ENCODE's third phase began, which lasted through 2016. This third phase would focus on strengthening the Data Coordination Center and Data Analysis, and develop the encodeproject.org website and matrix. Now, ENCODE data is hosted at the encodeproject.org website, where the data release policy has done away with the nine-month moratorium, meaning that data is required to be released as soon as possible, within a 24 hour window of their production (not enough time to stop data release or claim ownership over data). Researchers have no explicit rights over the data they produce, and publishing analysis is now a free-for-all. "External users," as the policy written on the encodeproject.org website described, "may freely download, analyze and publish results based on any ENCODE data without restrictions as soon as they are released."[127] The policy went on to encourage users to discuss data with producers,

---

[126] "ENCODE Consortium Data Release Policy Summary." *ENCODE Data Coordination Center at UCSC.* Accessed January 2017. http://genomebrowser.wustl.edu/ENCODE/terms.html

[127] "Data Use Policy" *Encode: the Encyclopedia of DNA Elements.* Accessed November 2016. https://www.encodeproject.org/about/data-use-policy/

but specified that this is optional. *Users* became a term for those who who may wish to publish on ENCODE data. The consortium as a whole, as well as individual members, will also publish results periodically. The policy also requested that researchers cite the most recent integrative consortium publication, and reference the specific data sets and laboratory(s),[128] indicating that publication is still important to the ENCODE consortium, but in an entirely different form from data release.

The National Human Genome Research Institute, the University of California Santa Cruz, and Stanford University, are all accredited as the creators of the encodeproject.org website.[129] Encodeproject.org is not intended to replace the University of California Santa Cruz genome browser or the other genome browsers which hold ENCODE data (Gene Expression Omnibus, Ensembl, and the International Human Epigenome Consortium), but to allow a new way for ENCODE data to be explored that is specific to ENCODE data. While the other browsers hold some ENCODE data, the encodeproject.org browser allows visitors to browse the entirety of ENCODE data.

This website would provide a way for ENCODE data to all be sorted and searched by anyone, and would allow multiple projects to integrate genomic data. The projects hosted on this website are, as of March 2017, ENCODE, the Genomics of Gene Regulation project, modENCODE, and the Roadmap Epigenomics Mapping Consortium, as well as other

---

[128] "Data Use Policy."

[129] The University of California Santa Cruz bioinformatics team had created one of the largest genome browsers to date, and was already affiliated with the ENCODE project. Stanford University housed J. Michael Cherry's research lab, and the Data Coordination Center for the ENCODE project.

epigenomics projects that are continually being added.[130] Metadata[131] was based off a combined effort from the Data Coordination Center and the Data Analysis Center, which allowed for the standardization of data. The Data Coordination Center and Data Analysis Center used a set of *ontologies*[132] to keep metadata controlled and coordinated, and created accessions to put metadata, particularly assay types, in context with one another. In the beginning of the ENCODE project, as mentioned in Chapter 1, ENCODE was formed out of a desire to create an ontology that built from or supplemented the Human Genome Project. Had it been like these ontologies, it would have been much more like a dictionary: a controlled set of vocabulary that provides a ground for a population to all be on the same page as one another linguistically. Instead, ENCODE became much more than that, and now used six different ontologies to keep its language consistent, making the ENCODE project much more encyclopedic, in that it built from dictionaries to bring vocabulary into three-dimensional life.

The encodeproject.org website is a "flexible platform that allows integration of genomic data from multiple projects," designed to make ENCODE data accessible and make all experiments reproducible.[133] It is a centralized source for all raw data, analysis data, methods, standards, and experimental metadata to be accessed,[134] and it also hosts the latest ENCODE data release policies and experimental protocols for those looking to replicate ENCODE data. The

---

[130]Cricket A. Sloan, et al. "ENCODE data at the ENCODE portal." *Nucleic acids research* 44, no. D1 (2016): D726-D732.

[131] These are tools or categories for defining and describing data.

[132] Each ontology (Uber Anatomy Ontology, Cell Ontology, Experimental Factor Ontology, Ontology for Biomedical Investigations, Chemical Entities of Biological Interest, and Sequence Ontology) is linked to at https://www.encodeproject.org/help/getting-started/#Ontologies. They each represent a consortium, some of which have overlapping membership with ENCODE, which each aim to standardize the vocabulary for different sets of biological concepts.

[133] Sloan et al. "ENCODE data at the ENCODE portal."

[134] Sloan et al. "ENCODE data at the ENCODE portal."

data visualization that it provides is specific to the data which the ENCODE project produces. Furthermore, it allows data to be put into a context of the software and pipelines used for their display; in addition to seeing how the data relates to one another on a genomics level, the software used for storing and displaying this data can be related on a software map that leads from one software type to another. Finally, this website introduced a feature called 'track hubs' which allow data to be visualized across genome browsers. In particular, this feature made ENCODE data much easier to relate back to other projects with data hosted by the University of California Santa Cruz genome browser or other genome browsers.[135]

While the ENCODE project established immediate rapid release upon the creation of encodeproject.org, the University of California Santa Cruz went in the other direction after the ENCODE project established its own browser. The University of California Santa Cruz genome browser launched a project in 2014 that accomplishes the exact opposite of rapid release. This project is called Genome Browser in a Box, and it allows personal, non-commercial use of the genome browser. Users could upload their own data to the browser and use the tools and structure of having a browser but did not have to release whatever they were working on to this public. This capacity of the University of California Santa Cruz browser lies in stark contrast to the ENCODE project's rapid release policy, which says that once data can be shared with the public, it has only 24 hours to do so. In this way, the ENCODE project's principles of rapid release are seen clearly in the encodeproject.org website, while the ENCODE project's previous locale on the University of California Santa Cruz genome browser shows no such principles present.

---

[135] Sloan et al. "ENCODE data at the ENCODE portal."

Developing its own genome browser has allowed the ENCODE project to establish its own practices and policies, and build ENCODE-specific software and pipelines. While the ENCODE project is still a project to produce data, far more resources need to go into the communication and framing of that data than did in the past. Thus, the third phase of ENCODE was focused on the Data Analysis Center and Data Coordination Center, who have produced and maintained encodeproject.org. These sub-organizations of the ENCODE project not only provided a home for the principles of rapid release to be held in the ENCODE project, but allowed all data to be fully framed by ENCODE, and for ENCODE purposes.

ENCODEPROJECT.ORG

The ENCODE encyclopedia, like the University of California Santa Cruz genome browser and the other channels on which ENCODE data is presented, is entirely customizable: there are at least thousands of different ways of looking at this mass of data, and more and data can be sorted in the most personalizable of ways. One of the main tasks of the Data Coordination Center is to be able to come up with more and more ways of sorting or categorizing data. To achieve this task, the Data Analysis Center and the Data Coordination Center work together to create pipelines, tools for the uniform processing of data to create more forms of "high-quality, consistent, and reproducible"[136] data. In creating more and more pipelines, more annotations for data, and more ways to sort and manage data, data is broken down into smaller segments, that are then arranged and sorted. The matrix, in this way, breaks the genome into many little pieces which an individual user can reshape or re-form in any way. Usership thus means taking the

---

[136] "Data Processing Pipelines." ENCODE: Encyclopedia of DNA elements. Encodeproject.org (Accessed November 2016)

ENCODE resource and reshaping it. Because the browser is so personalizable, and the data it displays so raw, the browser blends the ontological fact with the personal perspective.

Pipelines are software tools for maintaining this analysis process. They are continually updated and developed and they can be submitted by any consortium member or software developer. They are composed of discrete steps that can represent an algorithm, a software tool, or a file format manipulation that is applied to the primary data generated from an experimental assay. Essentially, they are a tool that allows researchers to break down their laboratory work into discrete algorithms, but also serve as mediators between different types of data by providing ways for data to be linked to one another and visualized. They are developed in correspondence with particular assays, using a platform called Github[137] which allows individuals to produce and share software with one another, for free if a project is open-source like ENCODE.

Github is used to make data repositories public for anyone who develops ENCODE-related software, and ENCODE prioritizes it for its ability to control for versions and updates publicly. This allows ENCODE data to be updated at any time, and for all software to be version-controlled. Users can move back in time through them, and use older versions, but they are continually updated to handle increased speed, efficiency, and sizes of data. Thus, as technologies for producing data are updated and improved, the technologies for representing and communicating this data publicly can keep up with its advances. Software development becomes the realization of ENCODE's goals to be an active agent in changing what is possible.

The use of github makes ENCODE data personalizable by allowing anyone to build their own software or collaborate with software developers to establish whatever mode of analysis or

137 "Github." Github. Accessed March 2017. https://github.com/

breakdown of data can be imagined. Software development becomes a completely open process through which anyone can follow, learn from, or participate. While it is completely open to the public, any ENCODE-affiliated software development is subject to the Data Use, Software, and Analysis Release Policies, which state that ENCODE requires the release of software for major ENCODE projects. In this way, the value of rapid release spills over from data release onto software development releases.

Many of these community-provided softwares emphasize their flexibility and breadth in their description, claiming their "wide variety of tools" (GATK), or "extensibility" (Java Treeview), or ability to perform on "millions of pairs" (King). Thus, these community-provided software tools extend the depth of ENCODE's reach. They enable ENCODE to connect more closely to other genetic and genomic studies, and enable them to do *more* in terms of high throughput data presentation. The openness with which these software tools can be submitted shows that anyone with an understanding of software development can look at the ENCODE genome as their canvas. The ENCODE project has become not only a project for creating an encyclopedia of DNA elements, but a project for creating the software for DNA elements. Software can undergo its own development and evolutionary processes, and the openness with which software is developed spills over into the genome by allowing the genome to be analyzed, broken down, and interpreted openly and individually.

In providing this personalizable take on the genome, the ENCODE project allows users to see whatever they wish to see. The consortium itself, then, can maintain a sense of being unbiased, non-analytical, apolitical: it is simply providing the medium on which users make their own work. Anyone can look at this browser and see whatever they see, and use it as a medium to

create something new, and communicate their thoughts and visions to others. The more the data can be analyzed and interpreted by users, the less analysis and interpretation the ENCODE project does of their own. Thus, to ENCODE, analysis means breakdown, and to analyze data is to make it open to new uses and interpretations.

The Human Genome Project's DNA sequence remained sturdy when framed by the Human Genome Project, but ENCODE has chiseled at it. Like a slab of shale, the ENCODE Data Analysis and Data Coordination centers break what appears as sturdy rock it into pieces that render it moldable like clay, analysis referring to the chiseling process, and coordination referring to the containment thereof. New projects, or users, can thus take whatever pieces of clay they desire and build their own forms from it. This is the new meaning for data to be published: rapid release removed data production from publication, and the moldable encyclopedia of DNA elements allows publication to refer to the molding and reshaping of data to create something new. Because *analysis* means *breakdown* to the ENCODE project, the ENCODE project's work is done when the genome is as moldable as possible.

While there may be a distinction between molding the genome in terms of looking at ENCODE data to shaping it for one's own analysis, and molding the genome physically on a living organism, the online genome may be the only tool for understanding the physical-form genome, and if the physical genome is *understood* to be moldable, then it is understood as moldable in either form. Increasing technologies that use genomic information may indicate that the genome is becoming moldable offline in the same way that ENCODE has made it moldable online. For example, technologies such as preimplantation genetic screening in which embryos are selected for insemination based on in-depth genetic analysis, have been referred to as

enabling 'flexible eugenics' because of the use of increasing genetic knowledge to change biological assets.[138] CRISPR/Cas-9 gene editing technologies which allow DNA to be spliced and edited on living organisms may indicate that the genome is increasingly seen as moldable by genetic scientists outside of the ENCODE project, beginning in 2012, the same time that ENCODE entered phase three.[139] The ENCODE project is increasingly understood as a tool for gene editing, and one lab from the Cold Spring Harbor Laboratory Press even already has published on the use of ENCODE data for RNA editing.[140]

These processes of data analysis and coordination refer to the breakdown of generated data into chiseled pieces, sorted and arranged for complete customizable and personalizable usership. Users, then, are the people who rebuild DNA from ENCODE's data. While the Human Genome Project produced a sturdy sequence of data, a dictionary, it lacked usership in this way. As the ENCODE project has brought DNA into an encyclopedia, it has made such data re-moldable in this way.

The Encyclopedia (of DNA Elements)

The 'encyclopedia' itself is a tab on the encodeproject.org website, presented as a 'matrix' which can be searched or arranged by a variety of categories. Data are presented as 'experiments,' and are sorted by a variety of continually updated metadata. What does the word *encyclopedia* mean to this project? There are not, nor are there ever intended to be, a printed

---

[138] Karen-Sue Taussig and Deborah Heath. "Flexible Eugenics: Technologies of the Self in the Age of Genetics." *Anthropologies of Modernity: Foucault, Governmentality, and Life Politics*. By Jonathan Xavier. Inda and Rayna Rapp. Malden, MA: Blackwell Pub., 2005. 194-212.

[139] Heidi Ledford. "CRISPR, the disruptor." *Nature:* News Feature. Vol. 522. Issue 7554. 03 June 2015. http://www.nature.com/news/crispr-the-disruptor-1.17673

[140] Park, Eddie, Brian Williams, Barbara J. Wold, and Ali Mortazavi. "RNA editing in the human ENCODE RNA-seq data." *Genome research* 22, no. 9 (2012): 1626-1633.

library of books or volumes of knowledge. So why does ENCODE describe itself as encyclopedic, and what is to be gained from the framing of this project as an encyclopedia? The encodeproject.org matrix is the only place in the ENCODE project that is referred to as an actual encyclopedia, but this matrix did not even exist until ENCODE was in its third phase.

'Encyclopedia' serves as a *master frame*[141] for the ENCODE project: it is a flexible but all-inclusive frame that encapsulates the entirety of the ENCODE project's activity. It also puts ENCODE in a context with other encyclopedia projects. Encyclopedias as a practice go back centuries, to Diderot's *Encyclopédie.* Early encyclopedias were much like maps in that they were imagined as circular or tree-like, despite their flat written form. They encompassed the entire circle of learning, or allowed for stems and branches on which knowledge blossomed, forming categories of knowledge and creating new forms of classification.[142] Early encyclopaedist Ephraim Chambers wrote that the key to the encyclopedia was the systematic ordering of such a wide range of knowledge: where the dictionary moved A-Z, the encyclopedia encapsulated the fuller circle of knowledge, moving cyclically instead of linearly.[143]

The Human Genome Project was a string of letters, A's, C's, T's, and G's, while ENCODE is far more complex and broad, hence the new questions committees of curation and coordination that emerged. Where the Human Genome Project was a dictionary, providing the words and where they fit in alphabetically, the ENCODE project is an encyclopedia, in that it provides in-depth explorations. It does not simply name the words, it explains them and allows

---

141 Snow et al. "Frame Alignment Processes, Micromobilization," 464-481.

142 Robert Darnton. "Philosophers Trim the Tree of Knowledge: The Epistemological Strategy of the *Encyclopédie*" *The great cat massacre: And other episodes in French cultural history*. Basic Books, 2009.

143 Darnton. "Philosophers Trim the Tree of Knowledge, 196.

people to use them, rather than simply read them. The framing of the project as an encyclopedia allows it to move outwards, further from the molecule DNA and into a broader living organism.

Over the years, the Data Coordination Center and Data Analysis Center have often been blended into one unified Data Coordination and Data Analysis Center. The Data Coordination Center's framing extended outwards to join with the Data Analysis Center. As analysis and coordination have merged, the difference between data that is produced and data that is used has subsequently become less distinguishable. In a recent funding opportunity announcement to expand the Data Analysis Center, the word 'encyclopedia' was referred to as a "compendium of candidate functional elements designed to enable exploration of the role of functional elements in disease mechanisms and basic biological processes."[144] The encyclopedia, in this way, is framed as a tool for understanding genetic 'disease,' and therefore bridges the gap between basic biology and applied medical science.

The meaning of the encyclopedia as a way of framing this project is that it moves the basic biological findings of the genome further and into a more embodied state by allowing this data not only to go deeper and give a fuller picture of the human genome, but bridge the gap between basic and applied biology. Data moves from production sites to Data Analysis sites to Data Coordination sites, and in doing so can swiftly, or rapidly, move from physical-form DNA to encyclopedic knowledge, which can be used, applied, and even published on, by the general public. In becoming three dimensional, it also can form the bits and pieces that make data into a more recognizable living organism, or a piece thereof.

---

[144] "RFA-HG-16-006" *National Institute of Health.* Accessed March 2017. https://grants.nih.gov/grants/guide/rfa-files/RFA-HG-16-006.html#_Section_I._Funding

In chapter two, I discussed how ENCODE is distanced from social applications by only allowing users to see tiny portions of a much larger genome. In breaking down the raw data into tiny pieces of annotation or code, the ENCODE project maintains this distancing by making each little piece unbounded to one another. Data points, in a matrix, do not resemble a human, but they are sourced from human and animal DNA. The more DNA is fragmented into data, the less it clearly represents a full human, making it more moldable to allow for individuals to form whatever image they wish to see on it. The online matrix becomes a medium for play with the human genome, a seemingly disembodied encyclopedia, a web of software on a screen, which seemingly lacks the repercussions of playing with a human body.

The human genome project elucidated a unified rock that was known as the genome, but the ENCODE project, in building upon that rock, has let it crumble into a giant slab of clay: moldable, but only if broken into pieces. It is no longer even sorted in the order which the code presents itself in DNA. In fact, now that it includes mouse genomes, data in the ENCODE encyclopedia does not have a species to stay bound to. The ENCODE project becomes a collection of pieces from many sources of shale, previously thought to be the sturdy rock of DNA sequences and of separate species, no longer distinguishable by type.

The ENCODE project's new definition of the words *encyclopedia*, *analysis*, and *users*, thus are indicate an *encyclopedia* is that which builds upon and breaks down the dictionary which the Human Genome Project provided, providing a much larger collection of data with which more can be done. While the Human Genome Project was two dimensional and flat, bound to order like a dictionary, ENCODE is three-dimensional and far more moldable and dynamic. *Analysis*, in this situation where data is released rather than published, becomes the

process of chiseling, and of breaking down the whole picture into little pieces that can be reshaped. *Users* become individuals who know how to look at the ENCODE clay and mold it into whatever forms they see fit.

If all genetic traits can be imaged *and* molded/reshaped through the ENCODE matrixes, the idea may pervade that physical-form DNA is equally moldable. In fact, Actor-Network Theory warns me that I should not try to distinguish between the physical-form DNA and the ENCODE matrix, because after all, the ENCODE project is a modality for imaging and understanding physical-form DNA. I am not, after all, looking at DNA, but at the way DNA is imagined and understood. If the image of DNA is to be molded and reshaped, whether or not DNA is actually molded and reshaped, it is understood as moldable and reshapable. The increasing views of the genome as a moldable source with which genes can be edited and reshaped with technologies such as CRISPR/Cas-9[145] gene editing that coincides with this process of moldability emerging in the ENCODE project indicates that this is part of a broader movement in genomics towards viewing the genome as unbounded and moldable.[146,147]

Gene editing is a highly controversial process, which many bioethicists, feminist, and sociologists of science have taken issue. The ENCODE project, however, has not faced the social critiques which gene editing has faced. Its biggest critiques, which will be discussed in the following chapter, are about ENCODE's claims rather than its applications. Francis Collins's metaphor of a three-story building of the ENCODE project would seem to be present here, where

---

[145] CRISPR/Cas-9 gene editing is a genetic technology that uses certain DNA segments, called Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR) to permanently edit genomes within organisms.

[146] Ironically, Jennifer Doudna, who was accredited as one of the inventors of CRISPR/Cas-9 gene editing technology, publicly called for a moratorium to distance the technology from its application.

[147] Nicholas Wade. "Scientists Seek Ban on Method of Editing the Human Genome." *New York Times:* Science. March 19, 2015.

ENCODE has affiliated with gene editing, but ENCODE still has evaded the social impacts of its findings because the ENCODE project is framed so openly. It is not a leader or director of the ENCODE project that has made ENCODE so open to interpretation and so moldable, but the overarching process of software development and encyclopedia production: the increasing number of customizable and open-sourced pipelines that make the ENCODE project open to anyone's interpretation.

Perhaps because the risks cannot be traced back to an individual researcher or moment, the risks may as well not exist to the institution.[148] This openness certainly does not favor any particular mode of breaking down the data set or analyzing ENCODE data, but does make ENCODE susceptible to the social currents that form around it. In being so open, ENCODE makes a distinct lack of indications towards applications of its findings. If the ENCODE project's only form of data regulation is to require the immediate relinquish of control over one's data through rapid release, the ENCODE project is not only giving up the ability to take a political stance, but it is encouraging whatever political stance is most pervasive surrounding the ENCODE project to overwhelm and dictate the applications of ENCODE data.

Conclusion

In summation, a dictionary is a much simpler presentation of data than an encyclopedia. The ENCODE project initially only had dictionary in mind as a model of what the ENCODE project would look like, but in turning two-dimensional DNA into a three-dimensional body of all DNA elements, it became an encyclopedia, and took on the characteristics that distinguish an

---

[148] An example of this phenomenon is depicted in:
Diane Vaughan. The Challenger launch decision: Risky technology, culture, and deviance at NASA. University of Chicago Press, 1997.

encyclopedia from a dictionary or ontology. A dictionary is far more attached to order, because if to disrupt the alphabet formation of the dictionary is to destroy it. An encyclopedia, on the other hand, can be re-ordered, and organized in a multiplicity of ways, such as chronologically or by category of data. In an effort to be a comprehensive encyclopedia, ENCODE establishes many forms of metadata, that allow the encyclopedia to be ordered in a variety of ways.

Because it is an encyclopedia and not a dictionary, it is still useful and valuable despite its ties to an essential unifying order. In becoming an encyclopedia, and establishing so much metadata, the genome becomes like clay, moldable and open to a variety of interpretations, rather than an ontology, which is a standardized and unified interpretation. It also establishes high risk, which become completely unregulated by rapid release policies, which have always been at odds with the moratorium, but in this current iteration make moratorium difficult, if possible at all. This shows that as DNA has moved beyond the data form, and into an encyclopedia, it has become unregulated and unrestricted. ENCODE, as a project which *develops new possibilities*, may be developing the possibility of a moldable genome, not only online, but in human bodies.

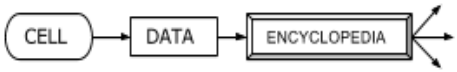
CHAPTER 4. ENCODE'S AUDIENCE.

Introduction

In chapter two, I explored some of the ways DNA is mediated from its physical form into the ENCODE matrix, establishing some of the framing techniques that allow for this process of translation from the body to the production sites or laboratories. In the third chapter, I explored some of the ways that data, taken from laboratories or otherwise, are translated into a matrix, or 'encyclopedia,' in cyberspace. For ENCODE's *users,* the project becomes uncontained here, and is now free to be molded to anyone's unrestricted using. However, there remains a multiplicity of other stances from which those who are not part of the ENCODE consortium look at and interact with the ENCODE project, from social media to classrooms to press releases to conferences and more. The purpose of this chapter is to begin to unpack some of these stances, or views from outside of the ENCODE frame, from those who are not ENCODE's direct users or consortium members, and thus to follow DNA, after being mediated by the ENCODE project's frames, beyond ENCODE.

In this chapter, some of the ENCODE project's fundamental frames are challenged or interactions with other frames, through processes of *polarization-vilification,* in which external actors frame themselves against ENCODE. I will show this by exploring the stances of four actors that are important to the ENCODE project: a recent publication, a partnership with the academic journal *Nature*, ENCODE's Twitter account, and Reddit, to see how they see the ENCODE project.

Each of these stances fills in a piece of what the ENCODE project looks like. They are all some of the possible viewpoints from which one could see the ENCODE project. A Latourian



Actor-Network Theory analysis should aim to inhabit each of these stances to look at the ENCODE project,[149] but for the sake of limiting this project, I will look at it from four stances which are central to the ENCODE project and to my telling of how the ENCODE projects frames have changed or remained the same since the ENCODE project's expansion.

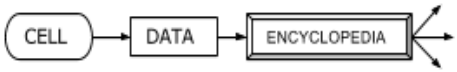
Each of these stances also ties into and completes some of the themes discussed in the previous chapters of *functional* DNA and of the moldable genome. In the previous chapters, I discussed what framed the ENCODE project, or kept it contained. In this chapter, I look at the challenging of some of its fundamental frames, or interactions with other frames, through processes of *polarization-vilification,*[150] in which external actors frame themselves against ENCODE. The ENCODE Reddit forum also provides an introduction to the future of the ENCODE project, the fourth phase, which was formally launched in March of 2017, in which the ENCODE project begins to show how it is affected by the currents around it, reframing ENCODE as not only an agent in changing the possibilities of the world around it, but a cautious subject of the changing possibilities of the world around it.

The many channels of ENCODE communication to the public include Twitter accounts, classrooms and lectures, Reddit, scientific publications by ENCODE members, scientific publications which cite ENCODE, videos on Youtube.com explaining ENCODE, conferences, press releases, the National Human Genome Research Institute website (genome.gov) or other websites affiliated with the National Institute of Health, science journalism, Facebook updates from an ENCODE Facebook account, and more. Each of these channels could provide a fruitful

---

149 Bruno Latour, *Aramis, or the Love of Technology,* Cambridge, MA: Harvard University Press, 1996.

150 McCaffrey and Keys. "Competitive framing processes in the abortion debate."

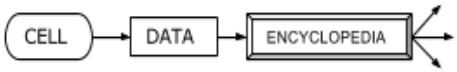


site to continue my frame analysis of the ENCODE project. They each hold their own characteristic and personas, as different paths through which data is translated. To keep this project limited in scope, I will here provide a small sampling of some of these channels, to see how the ENCODE project looks from a multiplicity of stances. I have been in the audience all along, but as a rogue member, who infiltrates spaces named as intended for the 'scientific community.' In doing so, I have shown how the ENCODE project looks from closer up. The ENCODE project holds a magnifying glass to human DNA molecules. If you stand close up to it, you can see the inner workings, the fragmentations that were discussed in the previous chapter. What, then, can you see from afar?

I will begin to answer this by sampling from these perspectives: the ENCODE consortium's most recent publication, the ENCODE project's partnership with the journal *Nature,* the ENCODE project Twitter account, and Reddit. Ultimately, they each have something to say about three controversial elements which have framed the ENCODE project since the 2003 launch and 2007 scale-up: ENCODE looks at *100%* of the genome, ENCODE is *developing new possibilities*, and ENCODE only looks at *functional* elements. The way these frames have shifted and changed indicates that ENCODE's future is powerful, ultimately to achieve a goal of fighting all human disease, a goal that is still in many ways undefined and unclear.

Publications

The ENCODE project publishes articles as a consortium every few years, which are distinct from the ENCODE project's more constant data release. While the previous chapter focused on data release, I will now look at the other side of the dualism between released and



published data that was established by exploring ENCODE's most recent published article. According to the encodeproject.org's expectations for citing ENCODE data, the most recent publication that has been published to the entire ENCODE consortium should be cited whenever ENCODE data is being cited.[151] These publications, beginning with a 2004 paper that was published in *Science*[152], can all be found on the National Human Genome Research Institute's webpage for the ENCODE project, along with ENCODE features and press releases.

Many of these publications are accredited to the entire ENCODE consortium, which removes individual authorship or responsibility, while other publications simply have a long list of authors named representing a wide variety of institutional affiliations. All of these publications are listed at the National Human Genome Research Institute's website. The first of these was published in 2004 in *Science* magazine, a prestigious scientific journal. It introduced the ENCODE project before the first phase had even been completed, and established, on a semi-public basis,[153] what ENCODE would be. In contrast to the 2003 ENCODE launch meeting, which established what the ENCODE project would look like to potential members of the ENCODE consortium, this publication established what the ENCODE project would look like to interested observers.

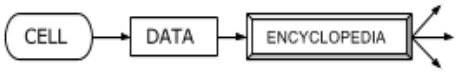
The most recent formal ENCODE publication listed by the National Human Genome Research Institute was published in 2014, although it is likely that there is another on the way due to the recent completion of ENCODE's third phase. This 2014 article is titled *Defining*

---

151 ENCODE DCC. "Data Use, Software, and Analysis Release Policies." *ENCODE: The Encyclopedia of DNA Elements.* Accessed December 2016.
https://www.encodeproject.org/about/data-use-policy/

152 ENCODE Project Consortium. "The ENCODE (ENCyclopedia of DNA elements) project." *Science* 306, no. 5696. 2004: 636-640.

153 This article is only available to those who subscribe to *Science.*



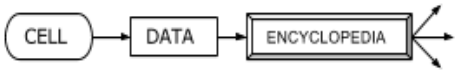
functional DNA elements in the human genome, and was published in the Proceedings of the National Academy of Sciences of the United States. This paper, although recent, reaffirms ENCODE's alignment with the Human Genome Project by opening with a mention of the completion of the Human Genome Project. It focuses on establishing what ENCODE means by the fifth and final anchoring frame that I discussed in chapter 2, *Function*: ENCODE will only look at functional DNA elements. It paper explains that function is a tricky word, which mostly relies on theories of evolution and tracing evolutionary constraint.[154]

According to this 2014 paper, the ENCODE consortium defines most of the genome as functional, but with varying degrees of evidence to proves function in each element. Only a very small portion of the genome codes for proteins, which has led to a belief that only a small portion of the genome has biochemical function, and that the rest is 'junk,' or DNA which simply takes up space but does not actually produce useful traits. This perspective that noncoding DNA was junk DNA was dispelled long before the ENCODE project began, but new definitions of junk versus functional DNA remained undefined. As discussed in Chapter 2, by defining the ENCODE project as a project which is only concerned with 'functional' DNA, ENCODE is anchored in a belief that there is a clear line between functions and nonfunctional DNA. Thus, the project would have to eventually define what this means. Here, in 2014, the ENCODE project has finally published a complex and nuanced definition. But why did it publish this definition now, 10+ years after using the word in the first place?

In 2012, the ENCODE consortium published a paper which claimed that 80% of the human genome's sequences had "biochemical function," using a vague definition of function that

---

[154] ENCODE Project Consortium. "The ENCODE (ENCyclopedia of DNA elements) project." *Science* 306, no. 5696 (2004): 636-640.



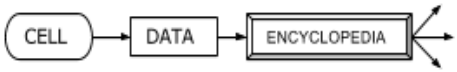
was based on whether a given sequence showed signs of replication, use by an enzyme, or evolutionary pressure.[155] This paper was met by a slew of controversy from other publications and science journalists, in particular, Dan Graur et al's *On the Immortality of Television Sets.*[156] Published in the journal of Genome Biology and Evolution, this paper scrutinized the concept of function in the human genome by saying that it was inconsistent, and that the definitions that the ENCODE project do provide of function are deeply flawed, sometimes fundamentally inaccurate. The titular claim, though certainly not the only bold argument of his, was that 'junk' DNA and 'garbage' DNA are two distinct concepts: 'junk' is that which you hold onto but will probably never use again, whereas 'garbage' is that which you get rid of because it causes a smell or takes up space or looks bad. Graur believed that much of what ENCODE described as 'functional' was actually junk, analogous to an abandoned television set left outside and unplugged for a long period of time. Even though the set is not going to get taken away any time soon, it is highly unlikely that anyone will ever watch a movie on it again, meaning that to claim that 80% of the genome has function is analogous to claiming that this television set is still functional.

Graur's paper not only attacked the semantics of the word function. He took his critiques of ENCODE further by claiming that the entire concept of an encyclopedia of human DNA elements was a foolish endeavor, and that it was informed by false notions of how evolution works. He framed his entire critique with the following epigraph:

---

[155]ENCODE Project Consortium. "An integrated encyclopedia of DNA elements in the human genome." *Nature* 489, no. 7414 (2012): 57-74.

[156] Dan Graur, Yichen Zheng, Nicholas Price, Ricardo BR Azevedo, Rebecca A. Zufall, and Eran Elhaik. "On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE." *Genome biology and evolution* 5, no. 3 (2013): 578-590.



> "I would be quite proud to have served on the committee that designed the E.coli genome. There is, however, no way that I would admit to serving on a committee that designed the human genome. Not even a university committee could botch something that badly" -David Penny (personal communication) (Graur et al 2013)

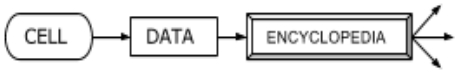
This epigraph places Graur's critique of the word function in a context of critiquing the entire concept of the ENCODE project. Graur's problem with the word 'function' was emblematic of a larger problem with the entire ENCODE project. The paper goes even further to claim that this is one of the faults of big data in general, and accuses the ENCODE project of overhyping. The paper ultimately finds that the ENCODE project has done more than its goal of producing data and releasing it sans analysis because ENCODE, in their 2012 publication, made a claim to how much of the genome is functional: "ENCODE's biggest scientific sin was not being satisfied with its role as data provider; it assumed the small-science role of interpreter of the data, thereby performing a kind of textual hermeneutics on a 3.5-billion-long DNA text."[157] He argues that if the ENCODE project was only concerned with providing data, it would not use analytical tools like 'function.'

Social movement theorists may refer to Graur's publication as an act of *frame debunking*. *Frame debunking* feeds off and exposes cultures of *polarization*, which is when an activity is framed by an us versus them mentality, and can be seen when one actor frames another as "corrupt, hypocritical, or a reprobate," which helps the first actor to be framed as a "moral agent fighting against evil."[158] Graur took one of the foundational frames of the ENCODE project and devalued it in a large and aggressive way, which in turn brought him publicity as a bearer of

---

[157] Graur, *On the Immortality of Television Sets.* 587.

[158] McCaffrey and Keys. "Competitive framing processes in the abortion debate." 44.



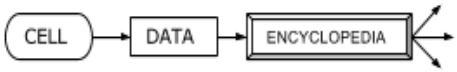
moral truth. Graur, in claiming that function analysis is not the role of a data repository, was saying that ENCODE had a conflict between two roles it inhabited of a project which released (not published) data and also a project which published a functional analysis. Perhaps ENCODE was breaking its own frame as a data repository by making an analytical claim, but this analytical claim had been one which the entire project was founded on.

ENCODE was claiming the presence of a binary between functional and nonfunctional DNA from the start: when the project launched it clearly stated that it would only focus on functional DNA elements.[159] The problem was that in 2013, this fundamental piece of the ENCODE framework became scrutinized. The problem was that this word, functional was the last of ENCODE's foundational frames to be fully defined, and in finally being *somewhat* defined by the 2012 publication, brought about a questioning of the ENCODE project as a whole.

In *Frame Analysis,* Goffman describes that a *negative experience* is when one is so focused on dismantling the framework of an experience, the experience of dismantling outweighs the original experience. It is is as if a new frame emerges out of the brokenness of the original frame, and contains the entire project, frames and all. Graur's critique is a perfect example of this, as he used one element of the ENCODE frame which he was steadfast on dismantling, to challenge and invalidate the entire ENCODE project. While Graur still saw and understood the inner workings of the project, he created a negative experience by using one critique of a framing technique to encapsulate and invalidate the entire project. As a result,

---

[159] NIH Center for Information Technology. *Encode Project Launch.*



ENCODE activity had to get to work on fixing that broken frame, and hence, released the 2014 publication *Defining Functional Elements in the Human Genome.*[160]

Soon after Graur's heated publication was released, there was an onslaught of publicity that was written about in a variety of biological reviews, such as W. Ford Doolitle's *Is junk DNA a bunk? A critique of ENCODE,*[161] published in the Proceedings of the National Academy of Sciences. New York Times science writer Carl Zimmer wrote an article titled *Is Most of Our DNA Garbage?*[162] and *Nature* published a variety of letters from scientists quickly weighing in opinions. Some of these articles sought to clarify what ENCODE was trying to say behind the word 'function,'[163] while others expanded on Graur's points.[164] Graur's team had, on their side, the accusation of sensationalization, critiques of big science, an a lofty literature review that allowed them to establish moral truths through accusations of inconsistencies and "statistical infraction."[165] The ENCODE project never formally fought back, aside from the 2014 paper which simply showed function to be more complicated than it was previously understood.[166] By claiming the matter to be more complicated than Graur had made it out to be, ENCODE refused to participate in Graur's *polarization-vilification*[167] practices, replacing Graur's binary

---

[160] Manolis Kellis, Barbara Wold, Michael P. Snyder, Bradley E. Bernstein, Anshul Kundaje, Georgi K. Marinov, Lucas D. Ward et al. "Defining functional DNA elements in the human genome." *Proceedings of the National Academy of Sciences* 111, no. 17 (2014): 6131-6138.

[161] W. Ford Doolittle. "Is junk DNA bunk? A critique of ENCODE." *Proceedings of the National Academy of Sciences* 110, no. 14 (2013): 5294-5300.

[162] Carl Zimmer. "Is most of our DNA garbage." *The New York Times Magazine. New York, NY: The New York Times Company* (2015).

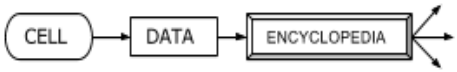
[163] Joseph R. Ecker, Wendy A. Bickmore, Inês Barroso, Jonathan K. Pritchard, Yoav Gilad, and Eran Segal. "Genomics: ENCODE explained." Nature 489, no. 7414 (2012): 52-55.

[164] Alexander F. Palazzo, and T. Ryan Gregory. "The case for junk DNA." *PLoS Genet* 10, no. 5 (2014): e1004351.

[165] Graur, *On the Immortality of Television Sets.* 583.

[166] Kellis et al. "Defining functional DNA elements"

[167] McCaffrey and Keys. "Competitive framing processes in the abortion debate."



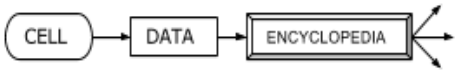
oppositions with a complex and nuanced look at the genome. In stead of joining into Graur's *polarization-vilification frame*, ENCODE maintained its stance that genomics is not a black and white issue.

After all of this controversy, it is clear why the ENCODE project's most recent publication would focus on forming a definition for functional DNA. The consequence, however, is that anyone looking to learn about the ENCODE project's most recent developments may begin with an article masked by this function debate. Function used to be a key that formed a small characteristic of the ENCODE frame, but this debate has brought it closer and closer to ENCODE's center-stage, and it is now somewhat of a *master frame,*[168] shielding ENCODE's other frames. This article may allow ENCODE to evade processes of vilification and polarization that were introduced by the debates Graur et al started, but it also means that the ENCODE project is looked at through a foggy lens of the *function* debate.

Nature | ENCODE

The 2012 paper in which ENCODE made the claim that 80% of the genome was functional was published in the prestigious scientific journal *Nature,* which ENCODE has a partnership with, and which has run multiple special volumes on the ENCODE project, all listed on a page called *Nature | ENCODE.* Although not always an openly accessible resource, the journal has an online page on the ENCODE that hosts many ENCODE-centric publications. The *Nature* page for ENCODE features a multimedia look at the ENCODE project, and is intended to be an introductory resource for learning about the ENCODE project. In addition to about six formal research papers, the *Nature* page shows two videos, a podcast on 'Big Science,'

[168] Snow et al "Frame Alignment Processes."



"companion papers" written by other affiliated genome research consortia, papers which use ENCODE, and papers that are otherwise affiliated with the ENCODE project. There are also a variety of news and comment papers that engage with social controversies around ENCODE.

On the *Nature | ENCODE* page, an animated video called "ENCODE: The story of you" explains how the concept of genetics has changed over the years, particularly putting the ENCODE project in a continuum with Mendelian genetics,[169] the Human Genome Project, and now, the Encyclopedia of DNA Elements. These past paradigms simplified the genome to a "purple and white issue,"[170] while ENCODE is large and complex and "a gloriously colorful thing."[171] As figure 4.1 shows, the ENCODE project is displayed as a hero of truth against these incorrect paradigms of genetics, and ultimately against disease, which is portrayed as a monster defeated by ENCODE. Here, *polarization-vilification* plays out in a different light: while the function debate incited by Dan Graur cast ENCODE as the villain against his truth, this video casts ENCODE as the hero against disease.

ENCODE: The story of you

Ever since a monk called Mendel started breeding pea plants we've been learning about our genomes. The latest chapter in our story is ENCODE; an ambitious project which aims to characterise all the functional element in the human genome. This animation shows how ENCODE is the next major step along this path.

Figure 4.1: The caption from *ENCODE: The story of you* and a still of the robot-like figure in which the ENCODE project is embodied. The image of the ENCODE project embodied is robotic, and shares a generally human-like form. However, it is portrayed as very large, and cannot even fit on the entire screen at once. After standing as pictured, the narrator says to "take a bow," and then it defeats a monster.[172]

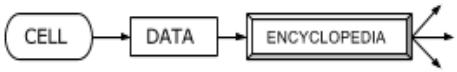
---

[169] Often understood as the first experimental genetics, this refers to Gregor Mendel's studies of inheritance on purple or white pea plants, in which he discovered that and how flower color is passed from parent to offspring, using a model of breeding purple or white flowers and tracing which offspring were then purple or white.

[170] Referencing Gregor Mendel's purple and white flowers.

[171] Nature video. *ENCODE: the Story of You*.

172 Nature | ENCODE. Accessed March 2017. http://www.nature.com/encode/#/threads



The podcast talks about the scale of ENCODE, and the challenges it provides by working with so many different scientists on one unified project, and over such a large time period. Speakers and writers discuss here how big science projects work and what it looks like to form consensus on such a large project. Editorials on *Nature* | *ENCODE* tend to focus on the size of the ENCODE project, such as a comment piece by ENCODE analysis coordinator Ewan Birney titled "Lessons for big-data projects,"[173] which lists many metrics of how the ENCODE project is organized, from numbers of experiments and participants to codes of conduct, to the cost of teleconferencing spent on the entire project, and an article called "ENCODE: The human encyclopaedia" which opens with the following:

> Ewan Birney would like to create a printout of all the genomic data that he and his collaborators have been collecting for the past five years as part of ENCODE, the Encyclopedia of DNA Elements. Finding a place to put it would be a challenge, however. Even if it contained 1,000 base pairs per square centimetre, the printout would stretch 16 metres high and at least 30 kilometres long.[174]

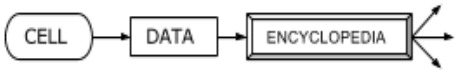
This opening statement frames the ENCODE project by its size. These *Nature* articles, which explain what the ENCODE project is focus on framing it by its size, emphasizing how daunting and unapproachable the project is. It falls into the category of 'big science,' and is too big to even wrap one's head around.[175] Framing the project in this way not only makes the ENCODE project seem unapproachable and daunting, but it shields out all other ways of looking at the ENCODE project. An onlooker, trying to learn about the ENCODE project by looking at *Nature* will first have to sift through their opinions on the controversy of 'big science' that frame ENCODE as too

---

[173] Ewan Birney. "The making of ENCODE: lessons for big-data projects." *Nature* 489, no. 7414 (2012): 49-51.

[174] Brendan Maher. "ENCODE: The human encyclopaedia." *Nature* News Feature. September 05 2012.

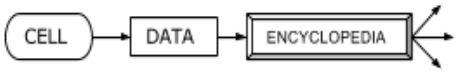
[175] Ewan Birney and Tejinder Virdee. Podcast: Science on a grand scale. *Nature* | *ENCODE.* Accessed April 2017.



big to visualize. They may then contextualize this size by watching *ENCODE: The Story of You*, where the project becomes a towering, heroic, robotic being.

This framing of ENCODE as 'big science' may be an offshoot or updated version of ENCODE's framing from the first scale-up from 1% to 100%. Now that it looks at the *whole* genome, it is a herculean project which has the size and strength to fight *all* of disease, with the power of all three billion nucleotides. In the previous chapter I discussed how the size of the human genome is so large that the DNA is fragmented into pieces. This page recasts ENCODE's size in a heroic light: ENCODE's size gives it the power to fight against the entire monstrosity of disease. In this way, the intimidation an onlooker might feel about the size of the ENCODE project is put to scale by the daunting size of the entirety of disease. This reframing of ENCODE's size indicates that however large ENCODE is, disease is larger, and if you are on the side of ending all disease, you will be on ENCODE's side, even if you are inclined against big science. This is another act of *polarization-vilification* because it puts ENCODE on one side and disease on another. However large disease is, ENCODE must be a comparable size. If disease is broken down into pieces, the frame of ENCODE as a large hero is lost, and if ENCODE is broken down into pieces, it loses its stability as a moral crusader.

At the same time, the *Nature* page makes the ENCODE project's size less intimidating by breaking down how 'big science' works, and portrays this size as what enables ENCODE to be a hero. In engaging with the size of 'big science' head-on, and breaking it down into metrics, the *Nature* page is able to use the framing of 'big science' to its advantage: instead of big science being a daunting and intimidating concept, it is the tool that we need to fight against the even bigger and intimidating monster that is all of human disease.



Twitter

In 2010, the National Human Genome Research Institute joined the social media website Twitter, with the account @genome_gov.[176] The following year, the University of California Santa Cruz Genome Browser joined Twitter with the account @GenomeBrowser,[177,178] and the ENCODE Data Coordination Center joined Twitter with @EncodeDCC. The year after that, in 2012, the ENCODE project joined Twitter with the account @ENCODE_NIH.[179] Various other affiliates such as genomics projects at the Sanger institute are on Twitter as well. Only the @genome_gov account is popular enough to have a Twitter validator. At the opposite end of the 'valid' spectrum is FauxENCODE, found at @FakeENCODE, an anonymously-run account that was created in 2013 (at the height of the controversy between ENCODE and Graur et al) and has not tweeted since 2013.

Twitter is a form of social media that allows users to make 140-character *tweets* that express anything from job opportunities to journal entries.[180,181] Sociologist Dhiraj Murthy wrote that Twitter can build from Goffman's theories of framing and performativity to expose the 'backstage' of people's lives: it questions the boundaries of public and private, by establishing public places that frequently are used to express private sentiments. Murthy argues that this puts

---

[176] Genome Gov, Twitter feed, accessed January 20, 2017. https://twitter.com/genome_gov

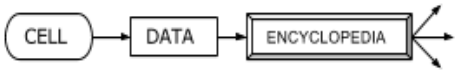
[177] UCSC Genome Browser, Twitter feed, accessed January 20, 2017. https://twitter.com/GenomeBrowser

[178] At the time, you may recall, this was the primary location for storing, sharing, and viewing ENCODE data.

[179] ENCODE Project, Twitter feed, accessed January 20, 2017. https://twitter.com/ENCODE_NIH

[180] Twitter users can also tweet at other users, which will lead the one being tweeted at to receive a notification, or they can reply to tweets. Twitter users can send each other direct, private messages. Tweets can feature hashtags which allow tweets to be linked together. Twitter users can search for tweets by account, hashtag, or any words used in tweets. Twitter has become an increasingly popular site for sociological research, particularly with the hashtags #Ferguson and #blacklivesmatter. If a Twitter is popular enough it will have a Twitter validator, a symbol which affirms for Twitter users that this account is not fake.

[181] Yarimar Bonilla and Jonathan Rosa. "#Ferguson: Digital protest, hashtag ethnography, and the racial politics of social media in the United States." American Ethnologist 42, no. 1 (2015): 4-17.



everybody in a 'global village,' a hybrid of public and private spaces that is both global and local.[182] Furthermore, Twitter ethnographies can show how people affiliate with one another. It may show endorsements, or alliances between organizations, particularly when organizations join Twitter. ENCODE inhabits Twitter in the form of a hashtag,[183] the @Encode_NIH account, the @EncodeDCC account, the fake ENCODE account at @FakeENCODE, and various posts that discuss controversies around the ENCODE project from outside actors.

In this section, I will explore the public activities of the @ENCODE_NIH account, to sample ENCODE's Twitter presence. I chose this site because it is both central to the ENCODE project, in that it is solely an ENCODE account, and because it touches the most external viewers of the ENCODE project by retweeting and following affiliated organizations, sharing articles and opportunities, and providing resources for researchers, policymakers, and educators. In this way, it is a central hub of ENCODE's connections to its audience.

The ENCODE project Twitter uses #ENCODE at the end of many tweets. These include data or software releases or genomics articles that are relevant to the ENCODE project.[184] The ENCODE Project frequently only interacts with other research projects' Twitter accounts, such as the Roadmap Epigenomics Mapping Center or ENCODE Data Coordination Center, or the Broad Institute in the United Kingdom.[185] The ENCODE project Twitter builds from the ENCODE encyclopedia by breaking up the data releases that the ENCODE encyclopedia hosts

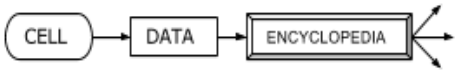
---

[182] Murthy, Dhiraj. *Twitter: Social Communication in the Twitter Age*. Cambridge: Polity, 2013.

[183] A hashtag is a way that a Twitter user can connect their tweet to a larger category of metadata, or a way that users can sort through the information they receive, by either posting with a # and an accompanying word, or by searching through a particular hashtag.

[184] See appendix I for examples.

[185] See appendix II for examples.



into even smaller pieces, hosted on an even more public medium. The processes discussed in the previous chapter, of ENCODE data fragmenting the genome, of ENCODE matrixes taking data out of the context of a universal order of the genome, are expanded onto an even deeper level, as data is released into the realm of social media, where each tweet has its own life, and can be separated entirely from the genome's sequence.

The ENCODE Twitter uses social media to connect and interact with others who frequently show pride and excitement for the ENCODE project. For example, when The Aiden Lab, a research laboratory, became an ENCODE mapping center, they tweeted their excitement and were subsequently retweeted by the ENCODE project account.[186] The ENCODE project follows 75 other Twitters,[187][188] from researchers to other research projects to other organizations. The ENCODE project does not only use Twitter to show updates and opportunities, but it shows a multiplicity of ENCODE activity- from affiliates to times ENCODE is mentioned by anyone at all. Furthermore, the ENCODE project has liked 22 posts. Shortly after first joining Twitter in 2014, ENCODE liked a series of posts participating in conversation around the function debate discussed in the previous section. Figure 4.2 A shows three tweets liked by the @ENCODE_NIH and B shows the conversation which one of these tweets was a part of, in which the Gruar versus ENCODE controversy was shown in a new light.

---

[186] Aiden Lab, Twitter post, February 2, 2017, 3:23 PM. https://twitter.com/theaidenlab/status/827296340975689728

[187] ENCODE project. Twitter feed.

[188] Any Twitter user may follow another Twitter user, or in the case of private Twitter accounts, request to follow another Twitter user, which will give them access through certain privacy settings such as the ability to direct message one another, and will lead the followed account to show up on the followers' homepage.

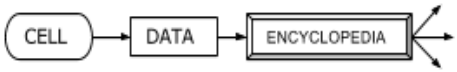


Figure 4.2. (A) shows a series of tweets liked by the ENCODE project in which two researchers and the staff from genomics and bioinformatics project *OpenHelix* participate in a debate denoted as #ENCODE. The "spanking" comment refers to OpenHelix accusing Dan Graur of giving ENCODE a "spanking." (B) contextualizes the first tweet by showing how R. Taylor Raborn filled in Jason H. Moore, PhD, on the debate he missed during his "twittercation." The link from R. Taylor Raborn leads to an article titled *The ENCODE Controversy and Professionalism in Science.*[189,190]

Twitter users tweeted, under the hashtag #ENCODE, that ENCODE's claims to 80% of the genome being functional questioned the entire theory of relativity[191] or that it could be used as a lesson in science blogging,[192] the importance of semantics,[193] while @Nature, the official account for the journal by the same name, tweeted that the debate "risks obscuring the real

---

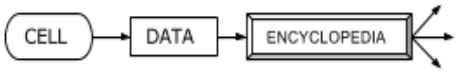
[189] Jalees Rehman. The ENCODE Controversy And Professionalism in Science. *The Next Generation: Divide and Differentiate.* February 24, 2013.

[190] ENCODE Project. Likes. Accessed March 2017.

[191] Jason Chin. Twitter post. April 6, 2013. 9:00 AM. https://twitter.com/infoecho/status/320566772619165697

[192] Marc RobinsonRechavi. Twitter post. March 6 2014, 6:55 AM. https://twitter.com/marc_rr/status/440863221977997312

[193] Marcela Preininger. Twitter Post. March 19 2013, 3:32 AM. https://twitter.com/mpreininger/status/313961282653024256



issue."[194] One Twitter account, anonymously run as @FakeEncode, challenged the question of professionalism by mocking the ENCODE project *and* making fun of Dan Graur (Figure 4.3).

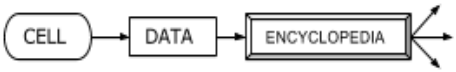
This anonymously-run Twitter account not only commented on the ridiculousness of Graur's article, as other participants in the debate had, but it commented on the ridiculousness of this debate taking place on Twitter to begin with. Overall, the ENCODE debate that took place on Twitter not only drew into question the professionalism of discourse in scientific publications, but brought the ENCODE project into a space where anyone could have a say on the project, as many scientists did.

Figure 4.3: @FakeEncode tweets two joke tweets about the ENCODE debate. One (top) was liked by two users and retweeted by one. The other (bottom) was only liked by the FauxEncode account itself.

@FakeEncode only has 77 followers,[195] but one of these was the @ENCODE_NIH blog. Its comments solidified the space of the ENCODE versus Graur debate on Twitter by creating an online platform centered around the debate, which dissipated when the debate faded out in 2014. In that it is completely anonymous, it opened up the idea that anyone could participate in the debate, while other researchers were tweeting about who could partake in the debate and how. It framed the ENCODE project as a controversy and a conversation. Perhaps this is the effect of ENCODE inviting the public in by joining Twitter: ENCODE was not only a knowledge base,

[194] Nature. Twitter Post. March 14 2013. 6:48 AM. https://twitter.com/nature/status/312198521908506624

[195] Fake ENCODE. Twitter Feed. April 2017. https://twitter.com/FakeEncode



but a spectacle, its 'spanking' and mocking a public warning of what science projects could and could not do.

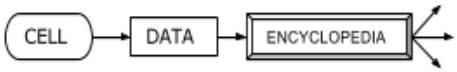
On the other hand, recall that ENCODE was being critiqued by Graur et al for veering too far on the analysis side of a data release/analysis dichotomy. Maybe inviting the public in by joining the ENCODE debate, and sharing ENCODE tweets, was a way of reaffirming ENCODE's lack of analysis, by leaving tweets as public offerings, as reusable and reshapable as data on the ENCODE matrix. The @ENCODE_NIH account did not partake in the debate any further than liking the three tweets in figure 4.2. Perhaps letting the debate take place without intervention was a way for ENCODE to bring data release to a new level, and reaffirm its side on the binary it created between release and analysis.

The medium of Twitter allows information and conversation to be broken up into small pieces, which can go used or unused, appropriated or ignored. This is not unlike the ENCODE matrix, which breaks DNA into small pieces to go used or unused, appropriated or ignored. By releasing ENCODE on Twitter, the ENCODE project took the principles of rapid release beyond the encyclopedia. By providing the space, through #ENCODE, for the debates around ENCODE to continue, ENCODE was putting to practice its commitment to letting anyone use ENCODE in any way they pleased.

Reddit

In the winter of 2017, the ENCODE project temporarily entered a new domain of social media: Reddit.[196] ENCODE researchers went live on Reddit through an account named

[196] Reddit is a form of social media that allows users to create forums, where they can post questions and begin conversations either anonymously or otherwise. There are "subreddits" which are categories within Reddit, and threads within these that are started by a moderator and joined by anyone in the public.



ENCODE_Project to let the public ask anything as part of a series of science themed *AMA*'s.[197] During *AMA*s, any question can be responded to by anyone, and users can vote on which questions the hosts should answer first. The hosts set a period of time which they will go 'live' to respond to users.

Shortly before this AMA, the National Institute of Health released a new round of ENCODE project funding to support five new centers "focused on using cutting edge techniques to characterize the candidate functional elements in healthy and diseased human cells."[198,199] The hosts were Nadav Ahituv from the University of California San Francisco, Elise Feingold and Mike Pazin who are two of the founders and leaders of ENCODE, Dan Gilchrist from the National Human Genome Research Institute's Computational Genomics and Data Science branch, and Yin Shen from the University of California San Francisco. They described the ENCODE project as an effort to unpack the *grammar* of the human genome, in contrast to the human Genome Project's *letters*, and said that the grammar and punctuation of DNA is hidden in the genome's 'dark matter.'[200]

This phrase, 'dark matter,' served as a way of reframing that which previously was referred to as 'junk DNA.' Instead of using the word 'junk' with all of its negative connotations, 'dark matter' implied that these DNA could be anything. Instead of saying that the ENCODE

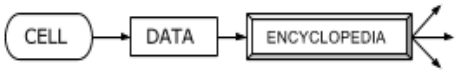
---

[197] AMA stands for "Ask Me Anything" and refers to a common practice on Reddit in which one will identify themselves- maybe they are a celebrity, or maybe they had a weird life experience- and create a thread on which anyone can post questions or comments. The person behind the AMA will stay online for a set duration of time in which they will immediately reply to comments or questions, meaning that conversations can be started. Conducting an AMA does not mean that any questions have to be answered, but that the host will make a concerted effort to answer as many as they can.

[198] ENCODE_Project. "Genome AMA." Reddit thread. February 9 2017. (Accessed March 2017). https://www.reddit.com/r/science/comments/5szvl4/science_ama_series_were_nih_and_ucsf_scientists/

[199] Note that, although this was 2017, and the paper describing the nuances of function came out in 2014, the word functional was still used as a master frame for the ENCODE project.

[200] ENCODE_Project. "Genome AMA."



project was focused on finding function in spaces where there was no function, the ENCODE project was framed as entering new territory. One of the questions directly addressed the question of functional versus junk DNA. User -Metacelsus- posted:

> I've just learned about repetitive DNA sequences (LTRs, LINEs, SINEs, etc.) in my biology class. Do you think they serve any important function, or are they just parasitic "garbage DNA"? What would happen if they were all removed?[201]

To which Nadav from the ENCODE project replied:

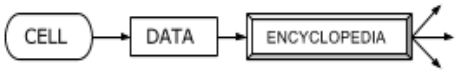
> Great question and one that my lab is actually very interested in and has active research on! With time and a lot of cool research, repeats are being found to have important functions in our genome. Many of them have been what's called "exapted." This is a term used in evolutionary biology to describe a trait that has been co-opted for a use other than the one for which natural selection originally built it. There are several cases where repeats have been found to turn into additional exons of existing genes, or gene regulatory elements that regulate other genes and change genome structure. Of note also, in the new phase of ENCODE, what we call affectionately call ENCODE phase 4, there is actually a computational group, led by Ting Wang from Washington University in St. Louis, who will specifically study the role of repeats in gene regulation. - Nadav[202]

This response reframed the question of "parasitic garbage DNA" in a few different ways. It detracted from the big-picture question of how much of the genome is functional, and instead focused on ENCODE's interesting findings in those areas. The ENCODE project was not claiming that large portions of garbage DNA did or did not exist, but that it would enter the 'dark matter' without deciding whether or not it was functional. It would no longer abide by a strict functional/non-functional dichotomy, but instead would look at the complexities and nuances of DNA. Much like how the ENCODE project blurred the boundaries between human and animal

---

[201] ENCODE_Project. "Genome AMA."

[202] ENCODE_Project. "Genome AMA."



studies when it entered phase 2 (as discussed in chapter 2), it would blur the boundaries between functional and nonfunctional DNA. Thus, instead of engaging with the notion of "parasitic garbage DNA" directly, he complicated and expanded on a matter which -Metacelsus- was, in Nadav's perspective, oversimplifying.

Furthermore, Yin from the ENCODE project went on to engage with questions of junk DNA twice. The first time, she simplified this into three points:

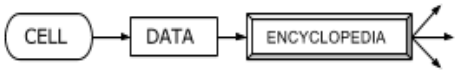
> The lessons we learned in the past ten years include: 1. There are millions of non-coding regulatory elements, a much bigger number than the protein coding sequences. 2. The regulatory elements are cell type specific and they are the major driving force for cellular identity. 3. A majority of the genetic variations associated with complex diseases are located in these regulatory elements, therefore mutations in these regions can play important roles in individual's susceptibility to diseases. -Yin[203]

This summary not only showed the importance of entering into non-coding 'dark matter' of the genome by describing how vast this region of the genome is, but framed this act of diving into genomic dark matter as a necessary way to understand complex diseases. Like how the *Nature* page claimed that anyone afraid of ENCODE's size would have to get over this in order to fight against disease, Yin's comment claims that anyone afraid of ENCODE's descent into genomic 'dark matter' or 'junk DNA' would also have to get over this in order to fight against disease. As seen in *Nature | ENCODE*, the *polarization* of ENCODE brought about by the *function* debate was thus shifted: rather than ENCODE against Dan Graur, ENCODE was against disease.

This was a new *master frame* of the ENCODE project: whatever else the project does or says could be justified if you look at ENCODE through a framework of defeating disease. This

---

[203] ENCODE_Project. "Genome AMA."



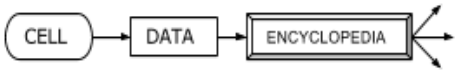
even changed some of the implications of the framing of ENCODE as an encyclopedia, as the encyclopedia referred to basic, unapplied knowledge, whereas ENCODE was now active, applied knowledge, on a mission to end disease, rather than a neutral resource. Now, the ENCODE project was certainly applied, only not to any specific or particular disease, but to an amorphous and gigantic body of diseases.

Yin and Nadav, who were from the University of California San Francisco, saw biochemical function as full of nuance, and in explaining this reframed the debate. Mike and Elise, who were ENCODE founders, had responses that were more faithfully aligned with the original five frames of the ENCODE project, in which functional versus nonfunctional DNA was a strict dichotomy. At one point, a Reddit user asked what classifies a sequence as "biologically relevant," to which Elise Feingold said that there are many different assays that can discover whether a sequence has function or not. This response equated what the Reddit user referred to as relevance to Feingold's definition of function, which reaffirmed the original ENCODE framework from 2003: ENCODE will only look at *functional* DNA.

Another Reddit user, Djabel1, directly asked about function in the debate over ENCODE's claim to 80% of the genome having function, to which Mike Pazin replied that the 80% finding was "an important first pass," but went on to explain some of the different ways in which one can define or study function will show different percentages. He elaborated that:

> An important part of ENCODE 4 will be its specific focus on examining candidate elements to determine whether, when, and where they function in important human cell types. This will be the task of the new ENCODE characterization centers, two of which Yin and Nadav will be directing at UCSF. -Mike[204]

[204] ENCODE_Project. "Genome AMA."



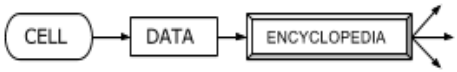
In this response, Mike Pazin continued to use the word function and continued to implicate it as a foundational characteristic of genetic elements being studied by the ENCODE project. He says that the future of the ENCODE project, as it enters phase 4, will be to specifically focus on function.

Perhaps Elise and Mike, as founding members of the ENCODE project, were less inclined to let go of the master frame of biochemical function in which the ENCODE project rested at its launch, because they saw the original purpose of *function* as an anchor for the ENCODE project. Newer members Yin and Nadav, on the other hand, were joining the project after it was heavily stigmatized by the ENCODE debate. The anchoring frames on which the ENCODE project was built were no longer the containers for the ENCODE project, and phase four was representative of this.

Phase four, which has only just begun in March of 2017, denotes a new uncharted territory for the ENCODE project. When asked what their biggest fears for the future were, Mike replied that his biggest fear was peoples' general lack of understanding as to how to interpret their genes, and failure to understand that the environment impacts the genome. This reflected the foundation of the ENCODE project as an encyclopedia to host complex understandings of the human genome. Nadav's response, on the other hand, was that "the obvious current scare is using CRISPR/Cas9 genome editing to make customized babies with traits parents want. In terms of realistic fears, I rate this as a 7 on a scale of 10."[205] This reflected fears for ENCODE's future, as a project susceptible to changing possibilities in the world around it.

---

[205] ENCODE_Project. "Genome AMA."



While Mike was still concerned with the matters on which the ENCODE project was founded, Nadav was engaged more critically with the changes and developments in technology that were more recent.[206,207] Recall that in the original ENCODE launch, ENCODE was framed as a project that would be an agent in changing what was technologically possible in the world around it. Nadav, however, is expressing fears of the technology around ENCODE developing further than he, as a leader of the ENCODE project, wants them to. In this way, the framing of ENCODE as an agent in changing the technological capacities of the world around it was shifted: as much as ENCODE would shape the world, it also had to be cautious of the development that was going on in the world around it.

Furthermore, one Reddit user expressed confusion that CRISPR/Cas9 was a scary technology, to which Nadav explained that it was scary because of how nuanced and complicated it was, making it potentially not suitable for practice:

> Many seemingly bad traits could also be considered to have good sides, or "silver linings." Will give you two quick examples: Beethoven and hearing loss and Van Gogh and schizophrenia. If you eliminate the "bad" traits from the human population, would you also eliminate the positive ones?'[208]

Like the *Nature* video which showed that ENCODE would fight disease by not approaching it from a "purple and white"[209] perspective as previous genetics paradigms have, Nadav is saying that to evade irresponsible uses of genetic technology, we must not approach genomics from a

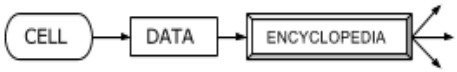
---

[206] CRISPR/Cas9 became increasingly more popular, funded, and saw a spike in patent applications around 2014.

[207] Heidi Ledford. "CRISPR, the disruptor." *Nature* 522, no. 7554 (2015): 20.

[208] ENCODE_Project. "Genome AMA."

[209] Nature video. *ENCODE: the Story of You*.



good versus bad traits perspective.[210] Perhaps these indicate that the future of ENCODE is to continue to do away with the binary oppositions on which it was founded (human/animal, computational/experimental, partial/whole, functional/nonfunctional). It must evade debates of functional/non-functional genetic elements by approaching the debate with nuance and complexity, and it must evade fears of the negative repercussions of gene editing by not viewing these from a good trait/bad trait dualism.

At the same time that her colleague was expressing his fears of CRISPR technology, Yin implicated the ENCODE project as a crucial factor in the future of gene editing technology:

> Scientists are working really hard to realize CRISPR tools for therapeutics for diseases…The purpose of ENCODE to understand how DNA sequences function in the non-coding part of the genome. Gaining that ability is essential in the precision medicine era when individual DNA sequences can be more easily obtained. Now the question is how to interpret millions of variants in each individual and which one will be the target. -Yin[211]

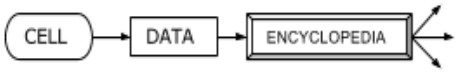
Although her colleague, Nadav, in the ENCODE project is afraid of CRISPR technology, Yin is implicating the ENCODE project as part of the development of CRISPR technology. ENCODE, as an agent in changing the technological capacities of the world around it, needs to engage in and enter the territories which it most fears, including gene editing technology.[212] Much like how ENCODE blurred the relationship between human and animal studies, despite its foundation as a

---

[210] Some of the consequences of a good traits/bad traits perspective have been written about in the following study, which looks at the impacts of *geneticization*, which is the practice of reducing human difference to genetic code: Sara Shostak, Peter Conrad, and Allan V. Horwitz. "Sequencing and Its Consequences: Path Dependence and the Relationships between Genetics and Medicalization 1." *American Journal of Sociology* 114, no. S1 (2008): S287-S316.

[211] ENCODE_Project. “Genome AMA.”

[212] CRISPR is only one of the applications of ENCODE technology that came up during the forum, and that there are many other applications of ENCODE in medicine and beyond. For example, Mike Pazin noted that pharmaceutical development can benefit from the project.

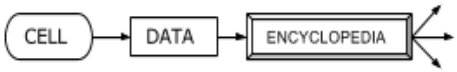


human-centric project, the ENCODE project would have to blur the relationship between being an agent in the changing technological capacities of the world, and being susceptible to the changing technological capacities of the outside world.

The ENCODE project was initially framed on a set of binary oppositions, such as human/nonhuman, or experimental/computational, functional/nonfunctional which have been dismantled or complicated, as the previous two chapters have traced. ENCODE remained an agent in changing what was possible in the world, but potentially in doing so opened up a pandora's box of gene editing technology. Or, as Yin explained, ENCODE can help gene editing technology be less risky. Ultimately, as this Reddit reflection on the function debate showed, the future of ENCODE looks less black and white, and rests less heavily on binary oppositions. But, like the *Nature* page indicated, the future of ENCODE is now framed by an overarching goal to defeat all of human disease. Overall, is the framing of ENCODE versus disease a new iteration of the same vilification, or is it a fresh start for ENCODE to be more committed than ever before to the complexities and nuances of the human genome?

ENCODE Against the World

Through scientific publications, Twitter, and Reddit, this chapter has explored how three of ENCODE's anchoring frames are challenged by its expansion upon these mediums. Scientific publications challenge the frame of functional DNA and expand on the frame of looking at 100% of the human genome. Twitter also challenged the frame of functional DNA and indirectly challenged the frame of looking at 100% of the human genome by critiquing 'Big Science,' but it reaffirmed the value of rapid release in the ENCODE frame. Reddit reaffirmed the frame of functional DNA and reaffirmed the frame of looking at 100%, but challenged the frame of



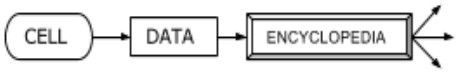
ENCODE as an agent in changing the capacities of the world around it, by discussing fears of the future of genome editing. While all three of these actors showed a need for nuance, and got rid of preexisting black and white binaries in the ENCODE frame, *Nature | ENCODE* and Reddit showed the emergence of a new master frame of the ENCODE project as the end to all human disease. Across all three platforms, a new master frame emerged, that strengthened and validated the ENCODE project by depicting its only true villain as disease.

In destabilizing some of the dualisms on which the ENCODE project was founded, it welcomed in the nuances of the relationship between genetics and society. Perhaps this is representative of a greater trend in which the dichotomy of the self versus the other is less recognizable in biology, the self being the genome and the other being the environment around it.[213] As founding leader Mike Pazin said when asked his greatest fear during the Reddit AMA, he worries that people do not understand how much the environment impacts our DNA. Another way of putting this would be to say that he fears a lack of understanding that the external world of the environment, and the internal world of DNA, are not dichotomous. In evoking this false dichotomy, he also indicates that nature versus nurture, and self versus other, may also be inapplicably dualisms in this new way of understanding genetic studies.

At the same time, the ENCODE project still maintains the self/other dichotomy when it comes to the dualism of ENCODE versus disease. This framing tactic, as discussed, emerges in times when other ENCODE frames are being debunked. Does this mean that the ENCODE project necessarily rests on binary oppositions in order to maintain validity? Without this opposition, will it be so uncontained that it perpetually manufactures *negative experience*? The

---

[213] Bruno Latour. *We have never been modern*. Harvard University Press, 2012.

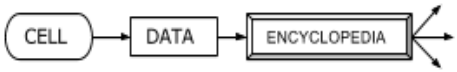


images against disease perpetuated in *ENCODE: The Story of You*[214] could indicate a violent new iteration of a dichotomy of self/other,[215] but how could the ENCODE project exist without this harsh duality? The ENCODE project has grown so big and so tall that it is embodied as the towering robot-like creature in *The Story of You,* and if it is not framed as a hero against some villain with some direction, it is terrifying to everyone. It could have chosen Dan Graur and his research team as its villain, but instead it chose human disease, a villain that is amorphous and undefined and maybe does not even deserve to die. If the ENCODE project's history is any predictor of its future, the framing of the entirety of human disease will soon have to be defined with clear parameters and openness to nuances, just as the framing of *functional* DNA did.

---

[214] Nature video. *ENCODE: the Story of You.*

[215] Anthropologist Emily Martin warns of a risk of violent dichotomies between the human self and disease in the following book:
Emily Martin. *Flexible bodies: Tracking immunity in American culture from the days of polio to the age of AIDS.* Beacon Press, 1994.



CHAPTER 5. CONCLUSION

FrankENCODE: Where Does It Roam?

This project has used a hybridized actor-network theory analysis and frame analysis to trace the movement of the ENCODE project as it has taken DNA from corporeal bodies to a cybernetic world. In the beginning, I established the five frames that the ENCODE founders set for the ENCODE project at its launch: it was *developing possibilities*, it was limited to *humans only* and *1%* of the genome, it was *computational and experimental*, and it was only concerned with *functional* elements. As the project progressed, each of these frames underwent various rekeyings, which my hybridized frame analysis and actor-network theory approach allowed me to follow by equipping me with the tools to trace the movement of a scientific project as both a social movement and an actor. The ENCODE project both creates frames and lives within them, and interacts with other actors that enter the ENCODE frames, such as the Bermuda Principles, *Nature,* or Reddit.

I have explored how each of the five anchoring frames has, in some ways, been challenged in the past fourteen years since the ENCODE launch. One of its roles in *developing possibilities* became about the ability to visualize the genome in new ways that reimagine the corporeal limitations of the human body, and allow the genome to be seen as a moldable substrate. The limitation of ENCODE to 1% of the human genome intentionally was dropped as ENCODE launched its second phase and became concerned with the entire genome. Along with this scale-up, the ENCODE project began to frame itself in defense of critiques or fears of big science, as seen in ENCODE's presence on *Nature* and Twitter. The meaning of *humans only*

shifted, and the ENCODE project now includes mouse genomes, and has offshoot projects which look at model species. Rather than this being a moment of ENCODE transgressing frames, looking at the meaning of human from ENCODE's perspective shows that the human and the mouse are not essentially different.

The practice of combining experimental and computational methodologies remained a central part of the project, and was even institutionalized through practices that allow ontological claims to be found at a crossroads of experimental and computational, making the two types of methodologies interdependent, rather than simply combined. Finally, the meaning of *functional* genomic sequences was increasingly questioned, resulting in a new and nuanced definition of function, and a polarized viewership resulting from the #ENCODE function debate. Along the way, ENCODE became framed as the hero against villainous disease, a new master frame which the ENCODE project relies upon when other frames are called into question.

Ultimately, none of these forms of movement could have been discovered without the integration of frame analysis, which provided tools and vocabulary for tracing the movement of a new scientific project. Though these theories are mostly drawn from social movement literature, the approach of Actor-Network Theory has allowed me to treat this scientific project as a social movement, by allowing me to follow ENCODE as an actor that builds, enters, and exceeds its own framing. Together they show me that while the ENCODE project is anchored, it is mobile, and it has moved and shifted throughout its life.

This project has told a story of the genome being fragmented into many pieces which could not be distinguishable between human or animal. As the ENCODE project developed its own website, or encyclopedia, this fragmentation was exacerbated to the point that data is now

presented like clay, from which users can rip off pieces, or build their own creations. This fragmentation of the body into encyclopedic form invites the human body into cyberspace, and allows humanity to exceed its previous capacities, and to be imagined in infinite new ways through a process I have titled the moldable genome. This provides new capacities for human creation and reproduction, and allows the future of human life to look nothing like it ever has before.

Much like the classical Frankenstein's monster, the ENCODE project was born out of a quest for knowledge, but in stretching the limits of human reproductive capacities, the founders of ENCODE created a new form of life, rogue and full of movement, and potentially imbedded with risks for an unpredictable future. The ENCODE project took the molecule of DNA and its accompanying genomic elements out of the human cell and gave it a new cybernetic life. How will this new iteration of the ENCODE project's life reflect back on the cell from which it was born?

The End of Human Disease?

According to postmodern theorist N. Katherine Hayles, processes of fragmentation have already rendered 'human' an inapplicable framework. The fragmentation of the body into pieces of information is a crucial part of how humans have become, as she argues, *posthuman*, by inviting what was once the human form into a cybernetic, virtual form, as a part of a trend of viewing human life as information.[216] If this is so, as soon as the ENCODE project expanded from 1% to 100%, it was a project which integrated the human biological organism and the

---

[216] Katherine Hayles. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago, IL: U of Chicago, 1999.

cybernetic computer simulation, and rendered the two indistinguishable, making the category of 'human' no longer relevant.

Furthermore, Hayles wrote that "in the posthuman, there are no essential differences or absolute demarcations between bodily existence and computer simulation, cybernetic mechanism and biological organism, robot teleology and human goals."[217] Thus, as the ENCODE project expanded to 100% and redefined *humans only,* it was only a matter of time before the computational and experimental became indistinguishable, and before questions of the relationship between robot teleology (or engineering) and human evolution came into question, as Graur's critique of the ENCODE project showed. Perhaps, by providing a space for the human body to be fragmented into bits of information, ontologies of the human body became inseparable from software engineering, creating a cybernetic-organic hybridized functionalism, making the social and the scientific indistinguishable, and depicting the human body as engineered.

Hayles argues that functionalism, or reducing experiences to the functions they serve towards a particular goal, is a part of this process, as a mediator between material life and codes of representation,[218] and that a necessary enabler of this process is the abandonment of the concept of 'human,' or at least an openness to redefinitions of the concept of 'human.' In creating life on a matrix, ENCODE made it possible to imagine new post-human life, which Hayles refers to as theoretical biology, or the construction of life as it could be, which allows form to triumph over matter. Thus, the reality which ENCODE has constructed within ENCODE frames do not

---

[217]Hayles. *How We Became Posthuman.*

[218] Hayles. How We Became Posthuman: 228.

remove DNA from a material world, but invite the cellular into the theoretical, blur theory and ontology, and allow the life sciences to also be the sciences of the ideal. What, then, are the ideals and theories that inform (or will inform) freshly molded genomes? With the moldable, or idealizable, genome, what will be built? The only answer the ENCODE project gives to this question is that ultimate ideal genomes will be free of disease.

Anthropologist Emily Martin wrote, in her book *Flexible Bodies,* about the impacts of an ideation of the body as something which can be molded and quickly adapted. She explored how concerns about the immune system with the AIDS epidemic led to increasing ideations of the human body as constrained to a limited 'tightrope' of evolution. The way people understood disease facilitated an imagination of the human body as at war with microbes, and for a person to be able to change their body meant that they had a leg up in that war. She wrote that this is a form of *neo-social-Darwinism* that sees a strong immune system as a way to scale and measure groups against one another. Having a flexible body became a value both inside and outside of medicine because it meant more freedom and adaptability in a context of heavy evolutionary constraint.[219] The moldable genome is not unlike the flexible body, able to be changed and edited, and able to give certain individuals a leg-up on evolutionary processes while others remain constrained. If ENCODE is human disease's biggest villain, will it (along with gene editing technology) lead to a molding of humans into the antithesis of known disease? Will this, subsequently, lead to a new iteration of *neo-social-Darwinism*?

If disease is ENCODE's sole remaining villain after breaking down the dichotomies of human/nonhuman, experimental/computational, and playing with partial/whole, it seems crucial

---

[219] Martin. *Flexible Bodies.*

that disease be defined and understood, but the ENCODE project has yet to indicate *which* human diseases it is against, only stating that it is against *all* disease. Although it has shifted the talk of a human genome away from being a black and white issue, the issue of human disease has gained a black and white framing. Such an undefined description of human disease raises the risk of eugenicist applications of idealized genomics. What dwells on the side of the idealized genome and what dwells on side of disease? Who decides what is disease and what is not? Will there be a home for deviant bodies in the moldable genome? Will there be space for neurologically diverse minds- for Beethoven's deafness and Van Gogh's schizophrenia?[220] What these examples expose is the way genetic diversity is deeply wrapped up with- sometimes indistinguishable from- identity. So what will this newly idealized, posthuman genome look like? The answer to this lies in the construction of disease.

Sociologists Anne Kerr and Sarah Cunningham-Berley wrote that the key to not being eugenicist is to maintain clearly on one side of certain social boundaries, particularly the boundary between neutral knowledge versus biased application, and choice versus coercion.[221] I argue that ENCODE has shown that knowledge, at least ENCODE's knowledge, is not neutral, in that the creation of an 'encyclopedia' became a project to bring about the end of disease, and because it facilitates the use of CRISPR/Cas-9 gene editing technology. Not taking a stand on gene editing *is* taking a neutral stand that allows surrounding social currents to inform the users of ENCODE, given the policy of rapid release that makes data immediately accessible to all sorts

---

[220] These are the responses from Nadav in the Reddit AMA discussed in chapter 4.

[221] Sarah Cunningham-Burley and Anne Kerr. "Defining the 'social:' towards an understanding of scientific and medical discourses on the social aspects of the new human genetics" *Sociological Perspectives on the New Genetics*. By Peter Conrad and Jonathan Gabe. Malden, MA: Blackwell, 1999. 149-170.

of users. Although the ENCODE project began as purely knowledge-based, it is now overwhelmingly framed by the vague concept of applications which the framing of ENCODE versus disease provides. Each of the boundaries set by ENCODE may have at first supported Kerr and Cunningham-Burley's idea that boundaries can prevent neutral knowledge from becoming coercive, but this essay has shown that even if boundaries are attainable, they are not sustainable, as rekeying and frame shifting processes have occurred and drawn into question ENCODE's all five of the ENCODE project's anchoring frames.

To frame the ENCODE project as the end of human disease is to take a stance, albeit vague, on the applications of ENCODE knowledge. What does a future of no disease look like? When a particular project focuses on a particular disease, they theoretically do more research than simply understanding the genetic elements to such a disease. If ENCODE is saying that it can end *all* of disease, as a genetics project, perhaps it is not reducing humans to genes, the way many accused the Human Genome Project,[222] but it is instead reducing all disease to genes, and neglecting the social components that shape human disease.

The Reddit conversation around gene editing in the fourth chapter points out that a moldable genome conjures up fears of the new eugenics (or neoeugenics[223]), in which fitness (as in physical traits perceived as evolutionarily advantageous, rather than as in physical strength) is determined by genetic assessments of traits,[224] or flexible eugenics, in which individuals live

---

[222] Horace Freedland Judson "A History of the Science and Technology Behind Gene Mapping" *The Code of Codes: Scientific and Social Issues in the Human Genome Project*. By Daniel J. Kevles and Leroy E. Hood. Cambridge, MA: Harvard UP, 1992. 3-36.

[223] Mukherjee. *The Gene: An intimate history*.

[224] Browner and Press. "The Normalization of Prenatal Diagnostic Screening."

with the knowledge that their biological assets are changeable,[225] whether or not they have access to this technology. The moldable genome, however, is not molded towards an increasing ideal, but in defense of a monstrous other. When the moldable genome reflects back onto human bodies, it provides the tools for reimagining complete genetic alteration.

Feminist scholar Anne Marie Balsamo puts medical imaging technology (a category in which ENCODE certainly fits as a technology for imaging genomes) in a context of plastic surgery, as the medical realization of bodily fantasies. She argues that "as the virtual body is deployed on a medium of information and encryption, the structural integrity of the material body as a bounded physical object is technologically deconstructed,"[226] making the body the realization of a "hacker's version of the American Dream."[227] Perhaps the new possibility which ENCODE opens up is an inverted future iteration of bodily modification, a posthuman plastic surgery, in which all human traits are up for debate, and the ideal final form is entirely undefined, prepared to shift and re-form based on whatever its villainous other throws at it.

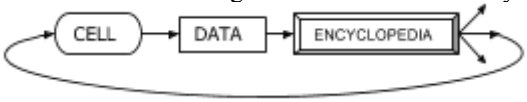
On the other hand, perhaps gene editing technology can be a source of empowerment. Donna Haraway has argued that we need to escape a mentality which assumes technology is against more vulnerable populations. In her "Cyborg Manifesto," she calls for a perspective that views women as living in tandem with technology, building new futures, empowered by technology to escape an oppressive past. If our future is as moldable as the ENCODE matrix depicts our genomes, we can rebuild our origins, rewrite the garden of Eden, and escape the

---

[225] Taussig and Heath. "Flexible Eugenics."

[226] Anne Marie Balsamo. Technologies of the Gendered Body: Reading cyborg women. Duke University Press, 1996: 131.

[227] Balsamo. Technologies of the Gendered Body: 131.



systemic oppression of women by building a clean slate to a new evolutionary tale.[228] The more radically the genome is chiseled apart, the deeper back we can go. To this ends, ENCODE has already taken the first few steps, by removing the essential order of the Human Genome Project and queering a duality of human versus nonhuman, and therefore entering a world free of essentialism of what any human needs to be.

As the ENCODE project moves forward, it must remember that it is not only an agent in shaping what is possible, but that it is also shaped *by* what is possible, in ways that are not limited to technological possibility, but the capacities for knowledge to be prescribed against atypical bodies. If the ENCODE project seeks to bring an end to disease, it must decide what disease is an is not, otherwise how is anyone to know if the traits that constitute our selves fall on the side of the hero or villain?

228 Haraway, Donna. "A manifesto for cyborgs: Science, technology, and socialist feminism in the 1980s." 214. *Feminism/postmodernism.* 1990: 190-233.

Appendix I.

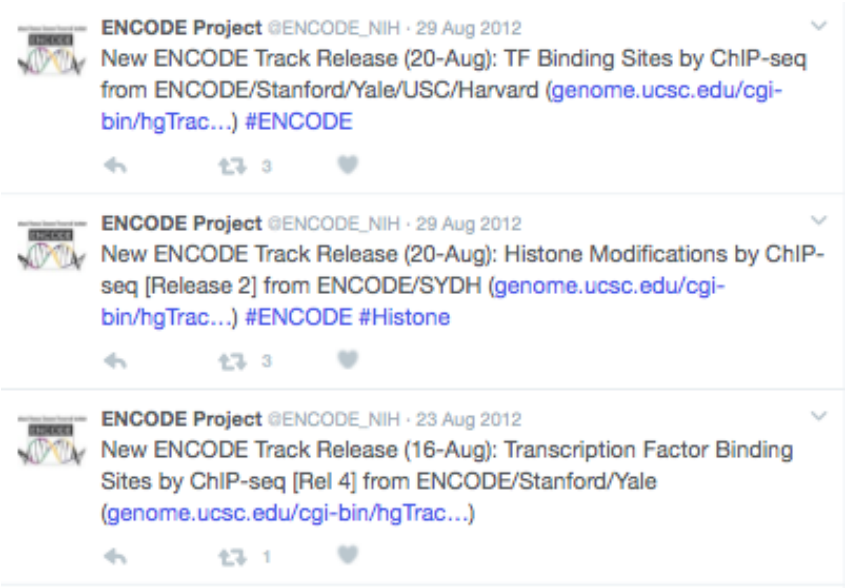


Figure 6.1.Three tweets from the ENCODE project Twitter in 2012 that show track releases (genome annotations) for the ENCODE project hosted on the University of California Santa Cruz Genome Browser. These show a common practice in the ENCODE Project twitter account to tweet quick summaries of software and data releases, all in the same format. [229]

[229] ENCODE project. Twitter feed, accessed January 20, 2017. https://twitter.com/ENCODE_NIH

B.

Figure 6.2. A series of seven consecutive tweets linking to articles about the ENCODE project. These show another common practice of the ENCODE twitter account, which was to tweet short topics and links to articles which use or mention ENCODE. It also shows the original use of the #ENCODE hashtag by the ENCODE Project, which contrasts the use of the #ENCODE hashtag in figure 4.2[230]

[230] ENCODE project. Twitter feed, accessed January 20, 2017. https://twitter.com/ENCODE_NIH

Appendix II.

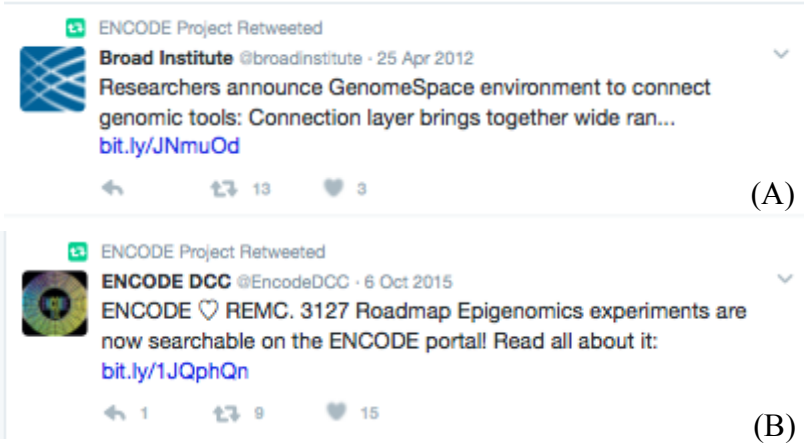


Figure 6.3. ENCODE project retweets other accounts. (A) ENCODE Project retweets the Broad Institute announcing a project to connect genomics researchers through cloud-based genomic analysis. (B) ENCODE ♡ REMC. The ENCODE project and the Roadmap Epigenomics project are not only affiliated, but in love, retweeted from the ENCODE Data Coordination Center.[231]

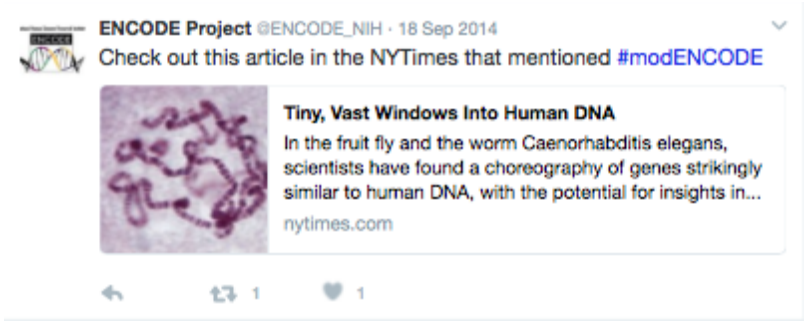


Figure 6.4. This article, shared by the ENCODE project Twitter account, talks about how the modENCODE project combines human and animal studies in a reciprocal way that displays the genome as, at once, tiny and vast.

[231] Both figures on this page are taken from:
ENCODE project. Twitter feed, accessed January 20, 2017. https://twitter.com/ENCODE_NIH

Works Cited.

"About the UCSC ENCODE pilot phase repository." *ENCODE Plot Project at UCSC.* Accessed February 2017.
genome.uwencode.org

"About the UCSC Genome Bioinformatics Site." *ENCODE UW.* Accessed February 2017.
genome.uwencode.org/ENCODE/pilot.html

Abu El-Haj, Nadia. *The Genealogical Science: The Search for Jewish Origins and the Politics of Epistemology.* Chicago: University of Chicago Press, 2012.

Almeling, Rene. "Selling Genes, Selling Gender: Egg Agencies, Sperm Banks, and the Medical Market in Genetic Material." *American Sociological Review.* 72, 2007: 319-40.

Aiden Lab, Twitter post, February 2, 2017, 3:23 PM.

Anonymous. "ENCODE debate revived online." *Nature: Research Highlights*. Vol. 509. Issue 7499. 05 May, 2014: 137.

Balsamo, Anne Marie. *Technologies of the gendered body: Reading cyborg women*. Duke University Press, 1996: 131.

Barnes, Barry, and John Dupré. *Genomes and What to Make of Them.* Chicago: U of Chicago, 2008.

Beck, Ulrich. *World Risk Society.* Cambridge, UK: Polity, 1999.

Benjamin, Ruha. "Designer and Discarded Genomes." e-flux Architecture. Accessed November 2016.

Benford, Robert D., and David A. Snow. "Framing Processes and Social Movements: An Overview and Assessment." *Annual Review of Sociology*, Vol. 26, 2000: 611-639

Berger, Peter L., and Thomas Luckmann. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge.* Anchor books ed. Garden City, N.Y.: Doubleday, 1967.

Bijker, Wiebe E., Thomas Parke. Hughes, and Trevor Pinch. "Society in the Making: The Study of Technology as a Tool for Sociological Analysis." *The Social Construction of*

Technological Systems: New Directions in the Sociology and History of Technology. Cambridge, MA: MIT, 1987.

----------"The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other." *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology.* Cambridge, MA: MIT, 1987.

---------- "The Evolution of Large Technological Systems." *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology.* Cambridge, MA: MIT, 1987.

Birney, Ewan. "The making of ENCODE: lessons for big-data projects." *Nature.* Issue 489, no. 7414, 2012: 49-51.

Bonilla, Yarimar, and Jonathan Rosa. "#Ferguson: Digital protest, hashtag ethnography, and the racial politics of social media in the United States." *American Ethnologist.* Issue 42, no. 1, 2015: 4-17.

Bourdieu, Pierre. “The Peculiar History of Scientific Reason.” *Sociological Forum* no. 6.1, 1991: 3–26.

---------- *Science of Science and Reflexivity.* Trans. Richard Nice. Chicago: U of Chicago, 2004.

----------*Outline of a Theory of Practice.* Cambridge, U.K.: Cambridge UP, 1977.

----------“The Specificity of the Scientific Field and the Social Conditions of the Progress of Reason.” *Social Science Information.* 14.6 ,1975: 19–47.

Brent, Michael R. "How does eukaryotic gene prediction work?" *Nature biotechnology.* 25, no. 8, 2007: 883.

Browner, Carole H., and Nancy Ann Press. “The Normalization of Prenatal Diagnostic Screening.” *Conceiving the New World Order: The Global Politics of Reproduction.* By Faye D. Ginsburg and Rayna R. Reiter. Berkeley: University of California Press, 1995. 307-21.

Brunet, Tyler DP, and W. Ford Doolittle. "Getting “function” right." *Proceedings of the National Academy of Sciences.* 111, no. 33, 2014: E3365-E3365.

Burr, Chandler. *A Separate Creation: The Search for the Biological Origins of Sexual Orientation.* New York: Hyperion, 1996.

Callon, Michel. "Some elements of a sociology of translation: domestication of the scallops and the fishermen of St. Brieuc Bay." *The Sociological Review.* 32, no. S1, 1984: 196-233.

Caplan, Arthur L. "Handle with Care: Race, Class, and Genetics." *Justice and the Human Genome Project.* By Timothy F. Murphy and Marc LappeÌ. Berkeley: U of California, 1994. 30-46.

Carninci, Piero. "Genomics: mice in the ENCODE spotlight." *Nature.* 515, no. 7527, 2014: 346-347.

Jason Chin. Twitter post. April 6, 2013. 9:00 AM.

Chadwick, Lisa Helbling. "The NIH roadmap epigenomics program data resource." *Future Medicine*: Vol. 4, No.3, June 2012: Pages 317-324.

Cussins, Chris. “Producing Reproduction: Techniques of Normalization and Naturalization in Infertility Clinics.” *Reproducing Reproduction: Kinship, Power, and Technological Innovation.* By Sarah Franklin and Helena Ragoné. Philadelphia: U of Pennsylvania Press, 1998. 66-99.

Cunningham-Burley, Sarah, and Anne Kerr. "Defining the ‘social:’ towards an understanding of scientific and medical discourses on the social aspects of the new human genetics" *Sociological Perspectives on the New Genetics.* By Peter Conrad and Jonathan Gabe. Malden, MA: Blackwell, 1999. 149-170.

Conrad, Peter, and Jonathan Gabe. *Sociological Perspectives on the New Genetics.* Malden, MA: Blackwell Publishers, 1999.

Cordner, A. “Strategic Science Translation And Environmental Controversies.” *Science, Technology & Human Values.* No. 40.6, 2015: 915–938.

Constant, Edward W., II. "The Social Locus of Technological Practice: Community, System, or Organization?" *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology.* By Wiebe E. Bijker, Thomas Parke. Hughes, and Trevor Pinch. Cambridge, MA: MIT, 1987. 223-43.

Cowan, Ruth Schwartz "Genetic Technology and Reproductive Choice: An Ethics for Autonomy." *The Code of Codes: Scientific and Social Issues in the Human Genome Project.* By Daniel J. Kevles and Leroy E. Hood. Cambridge, MA: Harvard UP, 1992. 224-263.

Cox, Susan M. and William McKellin. "'There's this thing in our family': predictive testing and the construction of risk for Huntington Disease" *Sociological Perspectives on the New Genetics.* By Peter Conrad and Jonathan Gabe. Malden, MA: Blackwell, 1999. 121-145.

Darnton, Robert. "Philosophers Trim the Tree of Knowledge: The Epistemological Strategy of the Encyclopédie" *The great cat massacre: And other episodes in French cultural history.* Basic Books, 2009.

DeSalle, Rob, and Michael Yudell. *Welcome to the Genome: A User's Guide to the Genetic Past, Present, and Future.* Hoboken, NJ: Wiley-Liss, 2005.

Djebali, Sarah, Carrie A. Davis, Angelika Merkel, Alex Dobin, Timo Lassmann, Ali Mortazavi, Andrea Tanzer et al. "Landscape of transcription in human cells." *Nature.* 489, no. 7414, 2012: 101-108.

Doolittle, W. Ford. "Is junk DNA bunk? A critique of ENCODE." *Proceedings of the National Academy of Sciences.* 110, no. 14, 2013: 5294-5300.

Downey, Gary Lee, and Joseph Dumit, eds. *Cyborgs & Citadels: anthropological interventions in emerging sciences and technologies.* New Mexico: School of American Research Press, 1997.

Duster, Troy. *Backdoor to Eugenics*. New York: Routledge, 1990.

——— "Buried Alive: The Concept of Race in Science." *The Chronicle Review.* The Chronicle of Higher Education, 14 Sept.

Ecker, Joseph R., Wendy A. Bickmore, Inês Barroso, Jonathan K. Pritchard, Yoav Gilad, and Eran Segal. "Genomics: ENCODE explained." *Nature* 489, no. 7414, 2012: 52-55.

"EDAC: Encode Data Analysis Center." *ENCODE.* Accessed March 2017.

EMBL-EBI. "Array Express- functional genomics data." *ArrayExpress.* Accessed March 2017. http://www.ebi.ac.uk/arrayexpress/

"ENCODE (Encyclopedia of DNA Elements) Facebook Page." Facebook.com Accessed January 2017.
Facebook.com/ENCODEProject/

ENCODE_Project. Genome AMA. Reddit thread. February 9 2017. Accessed March 2017. https://www.reddit.com/r/science/comments/5szvl4science_ama_series_were_nih_and_ucsf_scientists/

——— Likes (Twitter). Accessed March 2017.

———Twitter feed, accessed January 20, 2017.
https://twitter.com/ENCODE_NIH

ENCODE Data Coordination Center. "ENCODE Consortium Data Release Policy Summary." ENCODE Data Coordination Center at UCSC. Accessed January 2017.
http://genomebrowser.wustl.edu/ENCODE/terms.html

——— "Acknowledgements" *Encode: the Encyclopedia of DNA Element*s. Accessed November 2016.

——— "Contact." ENCODE: Encyclopedia of DNA elements. encodeproject.org. Accessed November 2016.

———"Data Use Policy" *Encode: the Encyclopedia of DNA Elements.* Accessed November 2016.

———"Data Use, Software, and Analysis Release Policies." *Encode: the Encyclopedia of DNA Elements.* Accessed December 2016.

——— "Data Processing Pipelines." ENCODE: Encyclopedia of DNA elements. encodeproject.org. Accessed November 2016.

——— "Encyclopedia." ENCODE: Encyclopedia of DNA elements. encodeproject.org. Accessed November 2016.

——— “Getting Started.” ENCODE: Encyclopedia of DNA elements. encodeproject.org. Accessed November 2016.

——— “News.” ENCODE: Encyclopedia of DNA elements. encodeproject.org. Accessed November 2016.

———“Project Overview.” *Encode: the Encyclopedia of DNA Elements.* Accessed December 2016.

——— “Publications using ENCODE data.” ENCODE: Encyclopedia of DNA elements. encodeproject.org. Accessed November 2016.

——— “Reference Epigenomes.” ENCODE: Encyclopedia of DNA elements. encodeproject.org. Accessed November 2016.

——— “Rest API.” ENCODE: Encyclopedia of DNA elements. encodeproject.org. Accessed November 2016.

——— “Tutorials & Workshops.” ENCODE: Encyclopedia of DNA elements. encodeproject.org. Accessed November 2016.

———“Ontologies.” *Encode: the Encyclopedia of DNA Elements.* Accessed December 2016. https://www.encodeproject.org/help/getting-started/#Ontologies

ENCODE Project Consortium. *ENCODE: Encyclopedia of DNA Elements*. Stanford University, 2016.

——— "An integrated encyclopedia of DNA elements in the human genome." *Nature* 489, no. 7414, 2012: 57-74.

——— "The ENCODE (ENCyclopedia of DNA elements) project." *Science* 306, no. 5696, 2004: 636-640.

———"Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." *Nature* 447, no. 7146, 2007: 799-816.

Ensembl. “ENCODE data in Ensembl” *e!ensembl.* published December 2016. Accessed Feb 2017.

——— “About the Ensembl Project” *e!ensembl*. Ensembl release 2017.

Ettorre, Elizabeth. *Reproductive Genetics, Gender, and the Body.* London: Routledge, 2002.

——— "Experts as 'storytellers' in Reproductive Genetics: Exploring Key Issues." *Sociological Perspectives on the New Genetics*. By Peter Conrad and Jonathan Gabe. Malden, MA: Blackwell, 1999. 37-57.

Fake ENCODE. Twitter Feed. https://twitter.com/FakeEncode. Accessed April 2017.

Feng Yue, Yong Cheng, Alessandra Breschi, Jeff Vierstra, Weisheng Wu, Tyrone Ryba, Richard Sandstrom et al. "A comparative encyclopedia of DNA elements in the mouse genome." *Nature* 515, no. 7527, 2014: 355-364.

Finkel, Elizabeth. *The Genome Generation*. Carlton, Vic.: Melbourne UP, 2012.

Gerstein, Mark B., Anshul Kundaje, Manoj Hariharan, Stephen G. Landt, Koon-Kiu Yan, Chao Cheng, Xinmeng Jasmine Mu et al. "Architecture of the human regulatory network derived from ENCODE data." *Nature* 489, no. 7414, 2012: 91-100.

Goffman, Erving. *Frame analysis: An essay on the organization of experience.* Harvard University Press, 1974.

Graur, Dan, Yichen Zheng, Nicholas Price, Ricardo BR Azevedo, Rebecca A. Zufall, and Eran Elhaik. "On the immortality of television sets: “function” in the human genome according to the evolution-free gospel of ENCODE." *Genome biology and evolution* 5, no. 3, 2013: 578-590.

“The Graveley Lab | Genetics and Genome Sciences | UConn Institute for Systems Genomics | UConn Health Center.” Accessed January 29, 2017. http://graveleylab.cam.uchc.edu/Graveley/.

Greil, Arthur L. “Infertile Bodies: Medicalization, Metaphor, and Agency.” *Infertility around the Globe: New Thinking on Childlessness, Gender and Reproductive Technologies.* By Marcia C. Inhorn and Frank Van Balen. Berkeley: University of California Press, 2002. 101-15.

Gieryn, Thomas F. *Cultural Boundaries of Science: Credibility on the Line.* Chicago: U of Chicago, 1999.

Gilbert, G. Nigel, and Michael Mulkay. *Opening Pandora's box: A sociological analysis of scientists' discourse.* CUP Archive, 1984.

Ginsburg, Faye D., and Rayna R. Rapp. *Conceiving the New World Order: The Global Politics of Reproduction*. Berkeley: University of California Press, 1995.

Guardian, The. "What the Encode project tells us about the human genome and 'junk DNA.'" Youtube video. Posted September 5, 2012.

Hallowell, Nina. "Doing the right thing: genetic risk and responsibility" *Sociological Perspectives on the New Genetics.* By Peter Conrad and Jonathan Gabe. Malden, MA: Blackwell, 1999. 98-120.

Haraway, Donna Jeanne. "A Cyborg Manifesto." *Simians, Cyborgs, and Women: The Reinvention of Nature*. New York: Routledge, 1991.

——— "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14.3, 1988.

——— *Modest_Witness@Second_Millennium.FemaleMan©_Meets_OncoMouseTM* Routledge: New York, 1997.

——— "Mice into Wormholes," in Downey, Gary Lee, and Joseph Dumit, eds. *Cyborgs & Citadels: anthropological interventions in emerging sciences and technologies.* New Mexico: School of American Research Press, 1997.

Harrow, Jennifer Adam Frankish, Jose M. Gonzalez, Electra Tapanari, Mark Diekhans, Felix Kokocinski, Bronwen L. Aken et al. "GENCODE: the reference human genome annotation for The ENCODE Project." *Genome research* 22, no. 9, 2012: 1760-1774.

Hayles, Katherine. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics.* Chicago, IL: U of Chicago, 1999.

Heath, Deborah, Rayna Rapp, and Karen-Sue Taussig. "Genetic Citizenship." *A Companion to the Anthropology of Politics*. Edited by David Nugent and Jan Vincent, 2008: 153-167.

Heath, D. "Locating Genetic Knowledge: Picturing Marfan Syndrome and Its Traveling Constituencies." *Science, Technology & Human Values* 23.1, 1998: 71-97. 4 Dec. 2016.

Helmreich, Stefan. "Abducting the Atlantic: How the Ocean Got Its Genome." *Alien Ocean: Anthropological Voyages in Microbial Seas*. Berkeley: U of California, 2009. 171-211.

Horn, David G. "Performing Criminal Anthropology: Science, Popular Wisdom, and the Body." *Anthropologies of Modernity: Foucault, Governmentality, and Life Politics*. By Jonathan Xavier Inda. Malden, MA: Blackwell Pub., 2005. 135-153.

Inda, Jonathan Xavier. "Analytics of the Modern: An Introduction." *Anthropologies of Modernity: Foucault, Governmentality, and Life Politics.* Malden, MA: Blackwell Pub., 2005. 1-19.

"International Human Epigenome Consortium." IHEC. Accessed January 17, 2017. ihec-epigenomes.org.

Jordanova, Ludmilla. "Interrogating the Concept of Reproduction in the Eighteenth Century." *Conceiving the New World Order: The Global Politics of Reproduction*. By Faye D. Ginsburg and Rayna R. Reiter. Berkeley: University of California Press, 1995. 369-86.

——— *Languages of Nature: Critical Essays on Science and Literature.* New Brunswick, NJ: Rutgers UP, 1986.

——— *Sexual Visions: Images of Gender in Science and Medicine between the Eighteenth and Twentieth Centuries*. New York: Harvester Wheatsheaf, 1989.

Judson, Horace Freedland "A History of the Science and Technology Behind Gene Mapping" *The Code of Codes: Scientific and Social Issues in the Human Genome Project*. By Daniel J. Kevles and Leroy E. Hood. Cambridge, MA: Harvard UP, 1992. 3-36.

Kerr, Anne, Sarah Cunningham-Burley and Amanda Amos. "The new genetics: professional's discursive boundaries." *Sociological Review*, 42, 2, 1997.

Keller, Evelyn Fox. *The century of the gene*. Harvard University Press, 2009.

———"Nature, Nurture, and the Human Genome Project." *The Code of Codes: Scientific and Social Issues in the Human Genome Project*. By Daniel J. Kevles and Leroy E. Hood. Cambridge, MA: Harvard UP, 1992. 281-299.

Kellis, Manolis, Barbara Wold, Michael P. Snyder, Bradley E. Bernstein, Anshul Kundaje, Georgi K. Marinov, Lucas D. Ward et al. "Defining functional DNA elements in the human genome." *Proceedings of the National Academy of Sciences* 111, no. 17, 2014: 6131-6138.

Kevles, Daniel J. "Out of Eugenics: The Historical Politics of the Human Genome." *The Code of Codes: Scientific and Social Issues in the Human Genome Project*. By Daniel J. Kevles and Leroy E. Hood. Cambridge, MA: Harvard UP, 1992. 3-36.

Kevles, Daniel J., and Leroy E. Hood. T*he Code of Codes: Scientific and Social Issues in the Human Genome Project*. Cambridge, Mass.: Harvard University Press, 1992.

Knorr-Cetina, K. *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard UP, 1999.

Laland, Kevin, Gregory A. Wray, and Hopi E. Hoekstra. "Does evolutionary theory need a rethink?." *Nature* 514, no. 7521, 2014: 161.

Latour, Bruno. *Aramis, or The Love of Technology*. Transl. Catherine Porter. Cambridge, MA: Harvard UP, 1996.

——— *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard UP, 1999.

——— *Reassembling the Social: An Introduction to Actor-Network Theory*. New York, NY: Oxford UP, 2005.

——— *We Have Never Been Modern*. Cambridge, MA: Harvard UP, 1991.

Latour, Bruno, and Steve Woolgar. *Laboratory Life: The Construction of Scientific Facts.* Princeton, N.J.: Princeton University Press, 1986.

Ledford, Heidi. “CRISPR, the disruptor.” *Nature: News Feature*. Vol. 522. Issue 7554. 03 June 2015.

Lublin, Nancy. *Pandora's Box: Feminism Confronts Reproductive Technology.* Lanham, MD: Rowman & Littlefield, 1998.

Martin, Emily. "The egg and the sperm: How science has constructed a romance based on stereotypical male-female roles." *Signs* 16.3. 1991: 485-501.

——— *Flexible Bodies: Tracking Immunity in American Culture from the Days of Polio to the Age of AIDS.* Boston: Beacon Press, 1994.

———*The Woman in the Body: A Cultural Analysis of Reproduction*. Boston: Beacon Press, 1987.

Martin, Paul A. "Genes as Drugs: The Social Shaping of Gene Therapy and the Reconstruction of Genetic Disease." *Sociological Perspectives on the New Genetics*. By Peter Conrad and Jonathan Gabe. Malden, MA: Blackwell, 1999. 15-33.

McCaffrey, Dawn and Jennifer Keys. "Competitive framing processes in the abortion debate: Polarization−vilification, frame saving, and frame debunking." *The Sociological Quarterly* 41, no. 1. 2000: 41-61.

Mcgonigle, Ian Vincent, and Ruha Benjamin. "The Molecularization of Identity: Science and Subjectivity in the 21st Century." *Genetics Research Genet*. Res. 98. 2016.

McGonigle, Ian V., and Lauren W. Herman. "Genetic citizenship: DNA testing and the Israeli Law of Return." *Journal of Law and the Biosciences.* 2: 469-478. 2015.

Mike Pazin. “How to Display and Download ENCODE data” *National Human Genome Research Institute.* Accessed April 2017. https://www.genome.gov/pages/research/encodeencode_display_and_download_pazin_13_05_08.pdf.

Morning, Ann Juanita. *The Nature of Race: How Scientists Think and Teach about Human Difference.* Berkeley: U of California, 2011.

Moore, Kelly, Daniel Lee Kleinman, David Hess, and Scott Frickel. "Science and neoliberal globalization: a political sociological approach." *Theory and Society* 40, no. 5. 2011: 505-532.

Laurence A. Moran. Five Things You Should Know if You Want to Participate in the Junk DNA Debate. *Sandwalk: Strolling with a skeptical biochemist.* Thursday July 04, 2013.

Mukherjee, Siddhartha. *The Gene: An Intimate History.* Simon and Schuster, 2016.

Murphy, Timothy F., and Marc Lappé. *Justice and the Human Genome Project.* Berkeley: U of California, 1994.

Murphy, Timothy F. "The Genome Project and the Meaning of Difference" *Justice and the Human Genome Project.* By Timothy F. Murphy and Marc LappeÌ. Berkeley: U of California, 1994: 1-13.

Murray S. Davis "Contemporary Sociology." *Contemporary Sociology,* vol. 4, no. 6, 1975: 599–603.

Murthy, Dhiraj. *Twitter: Social Communication in the Twitter Age*. Cambridge: Polity, 2013.

National Human Genome Research Institute. *"*NIH-DOE Guidelines for Access to Mapping and Sequencing Data and Material Resources*."* National Human Genome Research Institute. Last reviewed March 9, 2012. https://www.genome.gov/10000925/access-to-mapping-and-sequencing-resources/

——— ENCODE Project. "ENCODE Participants and Projects." *National Human Genome Research Institute.* Accessed January 2017. https://www.genome.gov/26525220/encode-participants-and-projects/

——— "ENCODE Project" Accessed January 26, 2017. https://www.genome.gov/10005107/encode-project/.

——— "ENCODE Project Background." *NIH National Human Genome Research Institute.* (Last Updated: May 21, 2012). Accessed March 2017. https://www.genome.gov/12513456/encode-project-background/

———"ENCODE/Roadmap Epigenomics Tutorial, Nov. 2012, ASHG." National Human Genome Research Institute (NHGRI). Accessed January 23, 2017. https://

www.genome.gov/27551933/ENCODERoadmap-Epigenomics-Tutorial-Nov-2012-ASHG.

———"ENCODE: The Encyclopedia of DNA Elements." Youtube Video. September 5, 2012.

———"ENCODE Workshop Summary - National Human Genome Research Institute (NHGRI)." Accessed January 26, 2017. https://www.genome.gov/10005115/encode-workshop-summary/.

———"NHGRI Rapid Data Release Policy." *National Human Genome Research Institute.* February 2003. https://www.genome.gov/page.cfm?pageID=10506376

———"The Human Genome Project Completion: Frequently Asked Questions." *National Human Genome Research Institute*, October 30 2010. Web. Accessed November 20, 2016.

——— "Using ENCODE Data to Interpret Disease-associated Genetic Variation- Mike Pazin," Filmed June 29-July 1, 2015, Published Aug 21, 2015. Accessed January 2017. https://www.youtube.com/watch?v=MKAv0pOJGUg

——— "Reaffirmation and Extension of NHGRI Rapid Data Release Policies: Large-scale Sequencing and Other Community Resource Projects." *National Human Genome Research Institute*. February 2003. Accessed February 2017. https://www.genome.gov/10506537/

——— "RFA-HG-16-006." Posted January 14, 2016. https://grants.nih.gov/grants/guide/rfa-files/RFA-HG-16-006.html

Nature. Twitter Post. March 14 2013. 6:48 AM.

Nature video. *ENCODE: the Story of You.* Youtube Video. Sept 10, 2012.

NCBI. "Gene Expression Omnibus." *National Center for Biotechnology Information.* Accessed March 2017. https://www.ncbi.nlm.nih.gov/geo/

NIH Center for Information Technology. *Encode Project Launch,* Videocast. March 7 2003, Washington DC. https://videocast.nih.gov/summary.asp?Live=2306&bhcp=1

NIH. “NIH to expand critical catalog for genomics research” *National Institute of Health*, February 2, 2017. https://www.nih.gov/news-events/news-releases/nih-expand-critical-catalog-genomics-research

Ni, Yunyun, Amelia Weber Hall, Anna Battenhouse, and Vishwanath R. Iyer. "Simultaneous SNP identification and assessment of allele-specific bias from ChIP-seq data." *BMC genetics* 13, no. 1. 2012: 46.

Nelkin, Dorothy and Lori Andrews. "DNA identification and surveillance creep" *Sociological Perspectives on the New Genetics.* By Peter Conrad and Jonathan Gabe. Malden, MA: Blackwell, 1999. 191-208.

Nelkin, Dorothy, and M. Lindee Susan. *The DNA Mystique: The Gene as a Cultural Icon*. New York: Freeman, 1995.

Nelkin, Dorothy "The Social Power of Genetic Information." *The Code of Codes: Scientific and Social Issues in the Human Genome Project*. By Daniel J. Kevles and Leroy E. Hood. Cambridge, MA: Harvard UP, 1992. 177-190.

Palsson, Gisli, and Paul Rabinow. "Iceland: the case of a national human genome project." *Anthropology today.* 15, no. 5, 1999: 14-18.

Palazzo, Alexander F., and T. Ryan Gregory. "The case for junk DNA." PLoS Genet 10, no. 5, 2014.

Panofsky, Aaron. *Misbehaving Science: Controversy and the Development of Behavior Genetics.* Chicago: University of Chicago, 2014.

Pazin, Mike, “Can an ‘Orchestra’ of Scientists Find the Hidden Music in your DNA?” from *TEDMED*, October 28, 2013.

Pazin, Mike, *How to Display and Download ENCODE Data.* NHGRI, NIH. Accessed January 2017.

Pei, Baikang, Cristina Sisu, Adam Frankish, Cédric Howald, Lukas Habegger, Xinmeng Jasmine Mu, Rachel Harte et al. "The GENCODE pseudogene resource." *Genome biology* 13, no. 9, 2012: R51.

Petryna, Adriana. "Science and Citizenship under Postsocialism." *Anthropologies of Modernity: Foucault, Governmentality, and Life Politics*. By Jonathan Xavier Inda. Malden, MA: Blackwell Pub., 2005. 158-177.

Petersen, Alan. "The new genetics and the politics of public health." *Critical Public Health* 8, no. 1, 1998: 59-71.

Point Mutation. Twitter post. February 21, 2013 at 2:59 PM.

Preininger, Marcela. Twitter Post. March 19 2013, 3:32 AM.

Rabinow, Paul. "Artificiality and Enlightenment: From Sociobiology to Biosociality." *Anthropologies of Modernity: Foucault, Governmentality, and Life Politics.* By Jonathan Xavier Inda. Malden, MA: Blackwell Pub., 2005. 181-193.

Rapp, Rayna. "Chromosomes and Communication: The Discourse of Genetic Counseling." *Medical Anthropology Quarterly.* 2.2, 1988: 143-57.

Roadmap Epigenomics. “jul2013.roadmapData.qc.” Google Docs. Accessed January 23, 2017.

———“REMC Standards and Guidelines for RNA-Sequencing DRAFT v3.0.” Roadmap Epigenomics, December 2, 2011.

———“Roadmap Epigenomics Project - Home.” NIH Roadmap Epigenomics Mapping Consortium. Accessed February 20, 2017. http://www.roadmapepigenomics.org/.

Roberts, Dorothy E. *Fatal Invention: How Science, Politics, and Big Business Re-create Race in the Twenty-first Century*. New York: New, 2011.

———*Killing the Black Body: Race, Reproduction, and the Meaning of Liberty.* New York: Pantheon, 1997.

Robinson Rechavi, Mark. Twitter post. March 6 2014, 6:55 AM.

Rothman, Barbara Katz. *Genetic Maps and Human Imaginations: The Limits of Science in Understanding Who We Are.* New York: Norton, 1998.

Sander, Jeffry D., and J. Keith Joung. "CRISPR-Cas systems for editing, regulating and targeting genomes." *Nature biotechnology* 32.4 (2014): 347-355.

Shakespeare, Tom. "'Losing the plot'? Medical and activist discourses of contemporary genetics and disability" *Sociological Perspectives on the New Genetics.* By Peter Conrad and Jonathan Gabe. Malden, MA: Blackwell, 1999. 171-190.

Shostak, Sara, Peter Conrad, and Allan V. Horwitz. "Sequencing and Its Consequences: Path Dependence and the Relationships between Genetics and Medicalization." *American Journal of Sociology* 114.S1, 2008: S287-316. JSTOR. Web. Mar. 2015.

Sloan, Cricket A. Esther T. Chan, Jean M. Davidson, Venkat S. Malladi, J. Seth Strattan, Benjamin C. Hitz, Idan Gabdank et al. "ENCODE data at the ENCODE portal." *Nucleic acids research* 44, no. D1, 2016: D726-D732.

Snow, David A., E. Burke Rochford Jr., Steven K. Worden and Robert D. Benford. "Frame Alignment Processes, Micromobilization, and Movement Participation" *American Sociological Review*, Vol. 51, No. 4, August 1986 p 464-481.

"Spanking #ENCODE." *OpenHelix*. Accessed March 2017. http://blog.openhelix.eu/?p=15398

Steinberg, Deborah Lynn. *Bodies in Glass: Genetics, Eugenics, Embryo Ethics.* Manchester: Manchester UP, 1997.

———*Genes and the Bioimaginary: Science, Spectacle, Culture.* Taylor & Francis, 2015.

TallBear, Kimberly. *Native American DNA: Tribal Belonging and the False Promise of Genetic Science*. Minneapolis, US: University of Minnesota Press, 2013.

Taussig, Karen-Sue, and Deborah Heath. "Flexible Eugenics: Technologies of the Self in the Age of Genetics." *Anthropologies of Modernity: Foucault, Governmentality, and Life Politics.* By Jonathan Xavier. Inda and Rayna Rapp. Malden, MA: Blackwell Pub., 2005. 194-212.

Taylor, Janelle S. "Confronting “culture” in medicine's “culture of no culture.” *Academic Medicine* 78, no. 6 (2003): 555-559.

UCSC Genome Browser, Twitter feed. Accessed January 20, 2017. https://twitter.com/GenomeBrowser

University of California Santa Cruz.“Conditions of Use - UCSC Genome Browser.” UCSC Genome Browser. Accessed January 23, 2017. http://genome.ucsc.edu/conditions.html.

———“Contact | Genomics Institute.” UCSC Genome Browser. Accessed January 23, 2017. https://ucscgenomics.soe.ucsc.edu/contact/.

———“UCSC Genome Browser Blog.” UCSC Genome Browser. Accessed January 23, 2017. http://genome.ucsc.edu/blog/.

———“UCSC Genome Browser: Citing Us.” UCSC Genome Browser. Accessed January 23, 2017. http://genome.ucsc.edu/cite.html.

———“UCSC Genome Browser: Contact Information.” UCSC Genome Browser. Accessed January 23, 2017. http://genome.ucsc.edu/contacts.html.

———“UCSC Genome Browser: Credits.” UCSC Genome Browser. Accessed January 23, 2017. http://genome.ucsc.edu/goldenPath/credits.html.

———“Genome Browser Gateway.” *UCSC Genome Browser*. Accessed January 2017. https://genome.ucsc.edu/cgi-bin/hgGateway

———“Genome Browser Licensing.” UCSC Genome Browser. Accessed January 23, 2017. http://genome.ucsc.edu/license/.

———“UCSC Genome Browser: News Archives.” UCSC Genome Browser. Accessed January 23, 2017. http://genome.ucsc.edu/goldenPath/newsarch.html.

———“UCSC Genome Browser: Staff.” UCSC Genome Browser. Accessed January 23, 2017. http://genome.ucsc.edu/staff.html.

———“October | 2014 | UCSC Genome Browser Blog.” WordPress. UCSC Genome Browser Blog. Accessed January 23, 2017.

http://genome.ucsc.edu/blog/2014/10/.

———"UCSC Genome Browser: Publications." UCSC Genome Browser. Accessed January 23, 2017. http://genome.ucsc.edu/goldenPath/pubs.html.

———"Track and Table Releases." Accessed January 23, 2017. http://genome.ucsc.edu/goldenPath/releaseLog.html.

Van Dijck, José. *The culture of connectivity: A critical history of social media.* Oxford University Press, 2013.

———*Manufacturing Babies and Public Consent: Debating the New Reproductive Technologies.* New York: New York UP, 1995.

Vaughan, Diane. *The Challenger launch decision: Risky technology, culture, and deviance at NASA.* University of Chicago Press, 1997.

Vernot, Benjamin, Andrew B. Stergachis, Matthew T. Maurano, Jeff Vierstra, Shane Neph, Robert E. Thurman, John A. Stamatoyannopoulos, and Joshua M. Akey. "Personal and population genomics of human regulatory variation." *Genome research* 22, no. 9, 2012: 1689-1697.

Watson, James D. "A Personal View of the Project" *The Code of Codes: Scientific and Social Issues in the Human Genome Project*. By Daniel J. Kevles and Leroy E. Hood. Cambridge, MA: Harvard UP, 1992. 164-175.

"Workshop Summary" *National Human Genome Research Institute,* Last reviewed: March 19, 2012.

Woolgar, Steve. "Reconstructing Man and Machine: A Note on Sociological Critiques of Cognitivism." *The Social Construction of Technological Systems: NewDirections in the Sociology and History of Technology.* By Wiebe E. Bijker, Thomas Parke. Hughes, and Trevor Pinch. Cambridge, MA: MIT, 1987. 311-28.

Yoxen, Edward. "Seeing with Sound: A Study of the Development of Medical Images." *The Social Construction of Technological Systems: New Directions in the Sociology and*

History of Technology. By Wiebe E. Bijker, Thomas Parke. Hughes, and Trevor Pinch. Cambridge, MA: MIT, 1987.

Yue, Feng, Yong Cheng, Alessandra Breschi, Jeff Vierstra, Weisheng Wu, Tyrone Ryba, Richard Sandstrom et al. "A comparative encyclopedia of DNA elements in the mouse genome." *Nature* 515, no. 7527, 2014: 355-364.

Zimmer, Carl. "Is most of our DNA garbage." *The New York Times Magazine.* New York, NY: The New York Times Company, 2015.