

**Compiled Public Comments on NIH Request
for Comments: Proposal to Update Data
Management of Genomic Summary Results
Under the NIH Genomic Data Sharing Policy**

Notice Number: NOT-OD-17-110

September 20, 2017 – December 12, 2017

Public Comments:

- 1 [Thomas E. Engells](#)
- 2 [Robert Jernigan, Iowa State University](#)
- 3 [Dean A Snyder, Johns Hopkins Genomics - CIDR](#)
- 4 [Jeffrey Barrett, Wellcome Trust Sanger Institute](#)
- 5 [John Quackenbush, Dana-Farber Cancer Institute](#)
- 6 [Abder Rahim Biad, AB Research](#)
- 7 [David Hunter, Harvard University](#)
- 8 [Bruce M Psaty, Univ of Washington](#)
- 9 [NHGRI-EBI GWAS Catalog, EMBL-EBI](#)
- 10 [Federation of American Societies for Experimental Biology \(FASEB\), Federation of American Societies for Experimental Biology \(FASEB\)](#)
- 11 [Paula Appel, student of Drexel University](#)
- 12 [Emmanuel Mignot, Stanford University](#)
- 13 [Matthew Parker, Sheffield Diagnostic Genetics Service](#)
- 14 [Prof Cathy Abbott, University of Edinburgh](#)
- 15 [Jason Merkin, KSQ THERAPEUTICS, INC.](#)
- 16 [Daniele Cassatella, University of Lausanne](#)
- 17 [Ian Quigley, Recursion Pharmaceuticals](#)
- 18 [Robert Castelo, Pompeu Fabra University, Barcelona, Spain](#)
- 19 [Frederick Roth, University of Toronto](#)
- 20 [Brooke Wolford, University of Michigan](#)
- 21 [Helen Naylor, Vanderbilt University Medical Center](#)
- 22 [Jessica Chong, University of Washington](#)
- 23 [Andrew McDavid, University of Rochester](#)
- 24 [JAMES Priest, Stanford](#)
- 25 [Eric Weitz, Not applicable](#)

- 26 [Keren Carss, University of Cambridge](#)
- 27 [Kevin Stachelek, University of Southern California](#)
- 28 [David von Schack, Pfizer Research and Deveelopment](#)
- 29 [Mette Nyegaard, Aarhus University](#)
- 30 [Amanda Spurdle, QIMR Berghofer MRI](#)
- 31 [Lori](#)
- 32 [Michael Hoffman, University of Toronto](#)
- 33 [Alicia Martin, Broad Institute & Massachusetts General Hospital](#)
- 34 [Christoffer Flensburg, Walter and Eliza Hall Institute](#)
- 35 [Karen Yu, BioMarin Pharmaceuticals Inc](#)
- 36 [Centre of Genomics and Policy, McGill University](#)
- 37 [Avinash Ramu, WUSTL](#)
- 38 [John Thompson, Claritas Genomics](#)
- 39 [Ian Dunham, European Bioinformatics Institute](#)
- 40 [J.A. Lathrop](#)
- 41 [Clare Turnbull, Institute of Cancer Research](#)
- 42 [Obi Griffith, McDonnell Genome Institute, Washington University](#)
- 43 [Gopala Tumuluri, UPMC](#)
- 44 [Dana Bis, University of Miami](#)
- 45 [Richard Bagnall, Centenary Institute](#)
- 46 [Prof Chris Semsarian, University of Sydney](#)
- 47 [Hugh French, Sydney Medical School](#)
- 48 [Kodama Yuichi, DDBJ Center, National Institute of Genetics](#)
- 49 [Natalie Nowak, Centenary Institute](#)
- 50 [Emma Singer, Centenary Institute of Cancer Medicine and Cell Biology](#)
- 51 [Laura Yeates, Centenary Institute](#)
- 52 [Jodie Ingles, University of Sydney](#)
- 53 [Michael Boehnke, University of Michigan Department of Biostatistics](#)

- 54 [Lea Davis, Vanderbilt University Medical Center](#)
- 55 [Daniel E. Weeks, University of Pittsburgh](#)
- 56 [Amelie Stein, University of Copenhagen](#)
- 57 [Chun-yu Liu, University of Illinois at Chicago](#)
- 58 [Bingshan Li, Vanderbilt University](#)
- 59 [Prescott Woodruff, UCSF](#)
- 60 [Ben Neale, Broad Institute](#)
- 61 [Brad Chapman, Harvard T.H. Chan School of Public Health](#)
- 62 [Ivan Limongelli, enGenome](#)
- 63 [Ryan Herringa, University of Wisconsin Madison](#)
- 64 [Stephen Faraone, SUNY Upstate Medical University](#)
- 65 [Vivek Ramaswamy, Mr.](#)
- 66 [Jamie L Wenke, Vanderbilt University Medical Center](#)
- 67 [Karol Estrada, Biogen](#)
- 68 [Rhonda Lassiter, Ambry Genetics](#)
- 69 [Chris Balak, TGen](#)
- 70 [Jack Kosmicki, Harvard Medical School](#)
- 71 [Warren Cheung, Children's Mercy Hospital](#)
- 72 [Sanjay Chandriani, BioMarin](#)
- 73 [Michael Squires, Law Student, S.J. Quinney College of Law](#)
- 74 [Miriam Rodrigues, Muscular Dystrophy Association of New Zealand](#)
- 75 [Janna Hutz, Eisai, Inc.](#)
- 76 [Samantha Barratt Ross, Centenary Institute](#)
- 77 [Joowook, Guy's Hospital](#)
- 78 [Michele Mattioni, Seven Bridges](#)
- 79 [Daniel MacArthur and Jessica Alfoldi, Broad Institute of MIT and Harvard](#)
- 80 [Ara Tahmassian, Harvard University](#)
- 81 [Larry Jacques, Sault Ste. Marie Tribe of Chippewa Indians](#)

- 82 [NHGRI-EBI GWAS Catalog, EMBL-EBI](#)
- 83 [Pamela Herd, University of Wisconsin-Madison](#)
- 84 [Daniel MacArthur and Jessica Alfoldi, Broad Institute of MIT and Harvard](#)
- 85 [Lisa Nichols, Council on Governmental Relations](#)
- 86 [James Lawlor, HudsonAlpha Institute for Biotechnology](#)
- 87 [Or Zuk, Hebrew University of Jerusalem](#)
- 88 [Allison Provost, Cohen Veterans Bioscience](#)
- 89 [Prof Mark Caulfield, Genomics England](#)
- 90 [National Society of Genetic Counselors, National Society of Genetic Counselors](#)
- 91 [Steven Buyske, Rutgers University](#)
- 92 [Denise A Dillard, Southcentral Foundation](#)
- 93 [Paul Flicek, European Molecular Biology Laboratory, European Bioinformatics Institute](#)
- 94 [Hae Kyung Im, University of Chicago](#)
- 95 [Jennifer Hall, American Heart Association](#)
- 96 [Joyce Tung, David Hinds, Adam Auton, Sarah Elson, 23andMe, Inc.](#)
- 97 [Maria Kannu, University Health Network](#)
- 98 [Erin Ware, University of Michigan](#)
- 99 [Steven Meyers, GISAID Initiative \(www.gisaid.org\)](#)
- 100 [Toby Bloom, New York Genome Center](#)
- 101 [Tejia Zhang, University of Utah](#)
- 102 [Goncalo Abecasis, University of Michigan](#)
- 103 [Nancy Cox, Vanderbilt University Medical Center](#)
- 104 [Degui Zhi, University of Texas health science center at Houston](#)
- 105 [Dr. Manuel A Rivas, Stanford University](#)
- 106 [Joon Yong An, University of California, San Francisco](#)
- 107 [Kenneth Mukamal, Beth Israel Deaconess](#)
- 108 [Mark McGary, Thyenmed](#)
- 109 [Liz Malerba, United South and Eastern Tribes Sovereignty Protection Fund](#)

Submission Date: 9/20/17

Name: Thomas E. Engells

Name of Organization: None

Type of Organization: Not Applicable

Role: Member of the Public

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

We will by design or default become the reference data set for genomic research. Other nations and stakeholders will harvest the benefit of the work without the requisite investment in the acquisition of the same.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

The decision to classify information as sensitive should be seconded by an independent and non-affiliated body of qualified assessors.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

A helpful tactic, but that decision (which datasets are sensitive) should be seconded by an independent and non-affiliated body of qualified assessors to include medical ethicists.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Again, we need to build in an accountability loop. A sample of users should be regularly reviewed to ensure that the data is being used responsibly. Genomic summary results could be a treasure trove of useful information, a building block in synthetic biology. We need to exercise caution and diligence in the handling of this information and take measures necessary to verify compliance with the stated approved uses.

Submission Date: 9/21/2017

Name: Robert Jernigan

Name of Organization: Iowa State University

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Risks are mostly unpredictable. Identification of individuals within the datasets is always possible, which needs to be guarded against. But, with proper precautions this would seem to be mostly unlikely. Clever terrorists might be able to devise population-specific biological and chemical agents, but this seems unlikely since only the most creative and knowledgeable (and evil) scientists would be able at present to do this. The benefits of this policy are clear. From the perspective of a researcher and user of data, access to these data will ensure that these data find broad application in many diverse research projects. The outcomes of this access are only partly predictable. One of the most obvious is permitting the development of background statistics to enable distinguishing real effects from those only marginally effective. Sufficient data would enable comparisons between different populations to more reliably assess the relative importance of individual SNPs. This distinction will have immediate impact on the practice of precision sequence-based medicine. And, without the release of this data the practice of precision medicine will be seriously hampered by being less reliable. In order to have reliable diagnostics such data must be used to make predictions, and reliable predictions require testing against large data sets. Already we are seeing interesting research projects utilizing the genome data accumulated and released by companies such as 23andme. However, unexpected applications of these data could be even more important. These data are critical to enable the development of a deeper understanding of evolution, which is critically important for understanding how to defeat many diverse existing and emerging pathogens. Quantifying risks and benefits are difficult. Both the risks and the benefits have some unknown outcomes. However, the apparent benefits must outweigh these uncertain risks in terms of benefits.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Making all of the data, even the sensitive data, available to appropriate individuals would seem to be generally useful. However, validating these users would require several steps.

These approvals might require the approval of the groups generating the data, the institutional approval of the requester, and final approval by a centralized committee. After initial experience these approvals might be reduced to the level of approvals required by the use of animal and patients. Ultimately modeling these approvals upon the IRB approval processes that have been so reliable would seem appropriate.

3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive

Putting the curation of this data in the hands of only the producers of the data is fraught with danger. Some centralized curation is likely necessary to ensure minimizing the risks for this data. Clearly some rules could be developed to reduce the risks within the submitted data, but ultimately minimizing risks requires final judgement that should be exercised by one data czar or one data committee. Naturally there would need to be regular channel for communicating between the data submitter and the centralized committee.

4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies

This would appear to be more or less acceptable, insofar as the risks had been minimized in the centralized screening of the data sets that were submitted. Perhaps some minimal validation of the institutions of the individuals accessing the data might be needed. (Neither North Koreans nor Iranians should apply.). These issues are extremely important for the future of biomedicine; thoughtful and broad-ranging considerations of how best to implement these policies are crucial. It is encouraging that these responses are being collected.

Submission Date: 9/21/2017

Name: Dean A Snyder

Name of Organization: Johns Hopkins Genomics - CIDR

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

It wasn't clear to me from reading the "Proposed Update to the NIH Genomic Data Sharing Policy's Access Model for Genomic Summary Results" if the following was addressed, but I feel that **every** discreet piece of released genomic data (that is, basically every individual file) should have a mandatory, standard header on it that:

- * can never be absent from the file*
- * documents the source of the file*
- * gives its legal status, that is, its restrictions on usage (such as a prohibition on reverse-engineering the data in support of disallowed activity)*
- * states that dishonest or malicious alteration of the header or the actual data is a criminal offense*
- * declares who is the most current, legitimate, authenticated, and responsible "owner"*
- * provides a chronologically ordered list of all previous "owners" for this physical file and its copies*

I think file-attached legal notification and provenance tracking is crucial for establishing accountability in the handling of this most personal of data.

Submission ID Date: 9/21/2017

Name: Jeffrey Barrett

Name of Organization: Wellcome Trust Sanger Institute

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

The benefits to providing unrestricted access to GWAS summary statistics are enormous. These files, when constructed appropriately, contain the vast majority of the most important information that can be gleaned from these studies. As such, making them available for secondary research has the potential to generate huge amounts of additional value from the (large) investments needed to generate the data. These can include better understanding of genetic relationships between human diseases, accelerated discovery of new associations, and more rapid translation of these studies into new therapeutics. In my view the risks are very close to zero. From the original publication of the Homer et al paper, the notion that a bad actor with access to a person's DNA sequence, and choose to use it to discover whether they had participated in a research study was far-fetched.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

This seems to be basically business as usual, which is fine.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

This would be an improvement to the current state of affairs, but even click-through mechanisms make it difficult to redistribute these summary statistics via other portals. I also don't believe such a click-through meaningfully reduces any risk (there aren't any), so would prefer to see it not there at all.

Submission ID Date: 9/25/2017

Name: John Quackenbush

Name of Organization: Dana-Farber Cancer Institute

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I am not sure what would be included in genomic summary results and so it is difficult to comment since a great deal would depend on the type of analysis used to derive them. Nevertheless, I would agree that there is minimal risk of any loss of privacy or re-identification of study participants with the release of such data.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I am not sure what personal data would constitute sensitive information that would be included in such summary data. I don't think institutions should be able to limit access due to considerations of their own scientific opportunity. This is a betrayal of the trust of the research subjects who consented because they wanted to advance our understanding of disease, not the careers of specific individuals.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

I would argue that summary data should be treated as non-confidential and that submitting institutions should make a case, to be reviewed by program staff at dbGaP, to determine whether such summary data should be designated as sensitive.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

My impression is that research subjects wish to see the data collected using their samples broadly used to advance searches for treatments for the diseases afflicting them. I think the NIH should do whatever it can to help assure broad use of those data in any manner in which there is minimal risk. I believe this is such an example.

Submission ID Date: 9/26/2017

Name: Abder Rahim Biad

Name of Organization: AB Research

Type of Organization: Not Applicable

Role: Member of the Public

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

A new field of research in human health is springing out that is opposite to everything we have known so far in healthcare: Biophysics engineering of the human body. Attention, care, and respect should be given the biophysicist engineer who knows well how DNA is energized and how it affects body functions is done with the accurate knowledge and approach. This medicine of the future will treat human diseases electrically and at the nanoscopic level; creating unheard-of change in the ailing person's condition.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

The conventional medical body tries to shy away from these innovators, who will be the sole authority to divulge the mysteries of the human body. Instead of treating symptoms, they will repair dysfunctional cells; fix their DNA and make them functional components in a massive network of a 100 trillion other components. New procedures and validation are needed for this new field of human health.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Allow the public to contribute to the database useful knowledge; only when such contributions are validated by an unbiased medical body. A person does not need to spend 10 years of his life, to be able to contribute to medical knowledge. I personally have been escorted out by security for approaching a university with revolutionary knowledge about restoring human life through Biophysics engineering of the human body. DNA expertise.

I am the author of Debugging Human DNA -- Lulu press -- found on Amazon.

Additional Comment (attachment):

Debugging Human DNA

Abder Rahim Biad

Medical innovations do not come exclusively from biology laboratories, but a horde of innovators and electrical engineers are working hard to simplify the medical profession. Success in this field always goes to medical doctors; innovators and engineers are always left in the shadow without recognition for the revolutionary inputs they contribute to human healing.

Outside medical circles, it is extremely difficult to claim any novelty in human health improvements. Debugging Human DNA is an area that has not been understood by medical science, because it is the domain of these explorers in the shadow who are also conducting their own research in unconventional ways.

The duty of this author is to offer valid scientific solutions, when others have failed to come up with answers that resolve our health problems. Conventional health providers are not equipped to understand a foreign domain, DNA functioning that can only be understood by electrical engineers. The human body is similar to an electrical apparatus with components that only operate when electrical energy flow through them. The various microscopic components inside the gigantic human machine and the over hundred trillion cells can only function electrically. Only a person with a good background in biophysics and electronics is

able to understand the human body's electrical functions; hence this book about Debugging Human DNA.

The medical community may suggest that the claim of knowing the secret to repairing human DNA outside conventional medical research is a fallacy; I will agree, if such claim was coming from an uninitiated; our doubt rises when we hear of any claims that are not validated by medical science, especially when the many discoveries in biophysics and nanotechnology are still at early stages and no significant ground has been broken in this domain of healing the human and prolonging life beyond the ordinary.

Human permanence is a topic that had puzzled medical scientists for centuries; DNA, the Foundation of Human Life has been also a piece of this puzzle. Most research on our genes is at early stages. The fact that the medical community has not broken any ground in finding its secret makes it challenging for an outsider to claim progress in this area of research, especially if he is not a health practitioner. Most of all, if I was a member of the medical community, I will be speaking the same language as the rest; I would be mentioning procedures such as this medication or that surgical procedure, but the truth is, I am an electronic engineer with background in physics and electronics, and only see the world from that standpoint.

Debugging Human DNA is an electrical engineer's domain, where exploration in this area is so daunting that it is almost impossible

to find practical resolutions in biology laboratories. The aspects of this innovation are so familiar to the biophysicist that it should be no surprise that such new discoveries would come from this body that is outside conventional biology laboratories.

Rest assured, I am not approaching this field from a conventional doctor's approach to human health but rather as a specialist of a field I know very well -- the electrical field. My conclusions are based on factual knowledge, confirmed by many years of experience as a system engineer. I am not coming to this enterprise in human healing as a conventional doctor but rather as an electrical expert with a background in electronic and psychoanalysis.

Psychoanalysis is indeed a factor in reaching these conclusions; I will explain what effect it has on Debugging Human DNA. These concepts are hard to imagine and are foreign to mainstream medical science; many of the conventional medical researchers do not understand the notion of healing the body electrically.

I have, nevertheless relied on resources from medical science to support and validate my finding on DNA restoration and I am certain about the information I present in this book; these are facts experienced by the author himself. I have tested many of the findings on myself and only mention those that I experienced through this method of transformation.

Debugging Human DNA is a new concept I am introducing to the reader for the first time; it is certain that without DNA functioning all over our body, decay will overtake the disconnected cells and causes the body to get sick and eventually die. This finding was discovered from research I conducted on myself. I have talked in my book: "Longevity" about Dead Weight Cells, (DWC), and their contribution to ailments and death and in this book: "Debugging Human DNA", I am presenting a continuum on the healing process by mastering the technique of debugging our own DNA; this is a finding no one was able to explain before the publication of this book.

The book came out of necessity to provide a valid approach to restoring human DNA to functionality. This is a new perspective on curing terminal diseases by correcting defective DNA. Suggesting this method of healing may astonish the uninitiated but it is an accurate assumption if we try to understand the inner working of DNA.

It is more likely that we can succeed modifying Human DNA by restoring it inside the human, than modifying it in new born babies and fetuses. Because even a new born baby with the best genes will age the same way as the rest; this suggest that modifying DNA inside an aged person, will improve his health and enhance the quality of his life and even extend it beyond the ordinary span.

As I stated, the brain can only be understood by a good background in physics and electronics. Health research is no more the domain of biology laboratories, especially when they have not provided satisfactory answers to terminal illnesses and premature death. The fact is -- this is an area of electrical engineering with all the apparatus of an electronics technician scanning test data on oscilloscope, to a computer programmer debugging the DNA strands, to an engineer interpreting the bits of information that our body emits.

Finding a solution to human suffering terminal illnesses is not the absolute responsibility of conventional medical researchers, but it is the concern of all who want to make life better for the terminally ill. Dying prematurely makes it an urgent duty to find a solution to premature death. Living longer is the dream of many but our failure to find a way out of death and the fear of this foe that is gobbling us by the millions, made us delusional to its destructive nature. Accepting our fate to die is the coward's way of admitting defeat to an enemy that plagues us all; we need to fight it as enemy number one and work hard at finding the solution to what is killing us.

Living beyond 100 is the dream of many, but to reach such goal one needs to debunk the mysteries of human life and adhere to the right healing system. Our cells harbor many secrets to a long and healthy life, especially working with our DNA and restoring it to functionality is the path to such venture. Debugging Human

DNA is a book that uncovers the secret to restoring our cells to health and curing the ailments they cause.

This may sound impossible outside medical labs and how can it be done by any individual without laboratory equipments? And how about knowledge of specific strands for specific human enhancement? The author argues that such goal is possible if we study the human body's electrical functions and for that I refer you to my books: "Longevity" and "Theory of Life"; in these books, I offer details about atoms and their energizing and explain how that affects the functions of our bodies.

Conventional medical research suggests that life prolonging cures can only be discovered in medical research laboratories and only when approved by such body that the use of such means are made available to the community; if these researchers could think outside the box and embrace the new revolutionary techniques of prolonging life beyond 100 years, they will find the answer to curing Cancer and Alzheimer's disease can only be accomplished through work with the body's electrical functions, on a nano-scopic scale.

The ignorance of conventional researchers is shown as many still mention neurotransmitters and cell carriers as if they were micro-organisms operating autonomously outside brain control. Suggesting the existence of a microscopic world inside our bodies with a will of its own to influence ailing neighboring cells; they

ignore the truth that life is nothing more than orderly atoms energized logically, performing the functions of replicating, strengthening and fighting intruders. Energizing DNA inside our cells, we cause the proper bonding of proteins inside them, to give us the shapes and forms we need, to exist as unique entities in a complex world.

One of the embarrassing things I heard in a lecture on Biophysics that protein is the CPU that processes DNA. How can a chemical compound have any logical functions when it is simply a form of matter with no logical faculties or channels within it to convey logical processes of our DNA?

The facts presented in this book are valid findings, based on scientific discoveries and are asserted by biophysics researchers. The discoveries can be explained with concrete details and can be tested for authenticity. Conventional Medical research has neglected an area that is outside biology; the new method to curing brain diseases and other bodily terminal diseases rely on the sanity of each cell's DNA; this is only possible when we have an understanding of electrical functions inside the human body.

Biophysics teaches us that the human being is mass and energy; energy impressed upon matter sustains life; this is a unique approach to curing terminal diseases and prolonging life. Selecting target DNA where discomfort is imminent, focusing neural energy on that cluster of DNA, will change the condition of

ailing cells and restore them to vitality. The answer to the dilemma of dealing with terminal diseases could not be found in biology labs, but rather going back to the basic concepts of life: the cell, and gaining an understanding of human life on a bio-physical level; it is there where the solution lies.

It takes an outsider to discover what conventional medicine could not, because explorers and innovators always operate outside the box. Medical researchers are waiting for the cure to come from laboratories and the promise of stem cells, but the cure is in the brain itself. To the cynics, if you have doubt, test the theory; you will find it is true; and once the facts are known, the new approach to healing terminal diseases will be the norm.

I am a system engineer who had studied electronics, psychology, and biophysics; I have discovered that the electronic theory applies to the brain and the body. Think of the human body as a car; it has mechanical parts that can be repaired by a mechanic or for that matter by a surgeon; it also has electronic circuitry that can only be fixed by an electronic technician – similarly by a neuro-scientist, educated in the new technique. The brain can only be understood by a good background in physics and electronics.

My understanding of the electronic theory and how electrical energy is generated and circulated in electronic systems has been a factor in this finding. As stated, the human body is similar

to an electronic system with components that only functions when proper electronic principles are followed.

You may say: How can it be done? Remember what I said earlier: This is a new approach; read what follows. Tai Chi, since ancient times, had mentioned how Chi (Energy) focused on ailing organs in the body, can heal and restore them to functionality. Tai Chi's teaching made me understand how neural energy flow throughout the entire body and is cause to restoring cells to health.

I have been studying human life throughout my entire existence; I have tried hard to find answers: to why we die? Why we get sick? Why can't we sustain life for more than 100 years? Living longer depends on our will to stay alive and on our knowledge of how to keep the body healthy and energized; to achieve such goal requires a lot experience and hard work.

Debugging Human DNA is a long process; it requires knowledge of electrical functions inside the body; knowledge of DNA strands; an understanding of computer programming will make it easier to understand DNA debugging, since DNA code is similar to computer binary code; they are all energized electrically.

With all DNA sequencing and the genome project DNA data base, there is still no broken ground on the inner workings of our DNA; the issue here is we do not know how these strands can be

assimilated into an aged body, to have any effect on improving human life. This book offers the right approach to using DNA to cure terminal diseases and prolong life.

Do not let all the DNA labeling jargons and the genome database discourage you from approaching this system of healing; there are a lot of DNA sketches out there put forth by researchers to explain the inner workings of our DNA; all the As, Cs, Gs and Ts are nothing more than labels we give to differentiate the various acids that DNA is. Inside the body DNA is a sensation; a variety of sensations our brain is able to interpret, sort, weed and store what is useful for later action. We are so used to this process that we do not pay attention to how these minute sensations affect our daily living. This book will teach you how to harness the secret of this *stat de chose*, so that you can perform the function of debugging your own DNA and restoring it to health.

The book's aim is to explain how DNA can be restored to help us fight the terminal and devastating diseases that are causing so many to leave this world prematurely. I assert the possibility to defy the odds and live longer is within reach; if we are open to new ideas. Throughout the ages, man was conditioned to die; we are accustomed to get ready to leave life when we reach a certain age.

When we become 60 years or older, we start making plans to depart this world and arrange for the way we want to be

disposed of. Many of us do not imagine outliving our keens; as soon as one spouse passes away, the other spouse will follow within few years or less. Sadness, misery, lack of purpose and mental illness cause us to want to leave life sooner. Living longer requires a lot of planning, hard work and the will to want to live.

The challenge is immense, because what the body used to do autonomously, you are going to do consciously and actively with regularity in order to change its condition from a sluggish dying organism to a healthy and long living body. Remember you can take the most potent serum deluded with all kind of magical ingredients, it will not do a thing to a dead cell; the only way out of this dilemma is to consciously and actively channel neural energy to dormant DNA inside dead cells, to bring them back to vitality and make them functional again.

Introduction

There is hope prolonging life beyond the usual, if we overlook the superficial knowledge we have about how genes work and what effect they have on the quality of our lives. There is, however, required knowledge about the role DNA plays in our mobility and practically anything we do in life. It is DNA that allows us to sort thoughts and libel sensations inside our brain; if you are unaware of this activity inside the human body, then it will be difficult to progress in this system of transforming human living.

What I am saying here is that there is required knowledge, without it, it is impossible to achieve success in this endeavor of healing the body through the brain. In order to change the stagnation in healthcare, the new knowledge is important for this technique to have effect on one's life. Most importantly, you need to opt out of the status-quo that our life's span is limited and one has to distract himself from the fear of dying until that day comes; that is the wrong attitude toward death and decay.

If you want to live longer you ought to take extraordinary measures beyond the usual, because there is no other alternative to a biological organism that is programmed to decay and fade away. Our autonomous body shuts down after so many years of use. This system deals with the working of the cell that is the building block of human life. It is certain the ailing body organs and parts can be cured within the human body with minimal surgical intrusion. To make this approach a reality, we need to focus our attention on the working parts of a human cell and gain an understanding of how DNA is affecting our cells' health.

It should be no surprise that a book of this kind would mention alternatives to tackling the hard to cure diseases. And the reason it matters, of course, is that those who have no time or intellect to see an alternate path through the maze of confusion in all the research we bring to the table and yet we have no idea what works and what does not. The sense of urgency is to have the power to dispel the lack of enthusiasm for attacking aging, by

reinforcing the suggestion that the medical community should look elsewhere for answers to curing the devastating diseases that are plaguing us and taking us away from life by the millions.

The techniques in this book are unique in the way they will cure terminal diseases and make the individual human live longer. I also address other medical issues, relying solely on my extensive research in human behaviors and my curiosity to understand our purpose for living. I learned that mental illness is not a static phenomenon but rather a dynamic condition; given the proper knowledge of self and the right approach to healing, one can indeed solve mental problems and lead a normal life. You may say: why is the cognitive is relevant to health? It is relevant because the cognitive affect the sensory inside the entire the body; we function with both cognitive and sensory stimulation.

The new system to cure terminal diseases and prolong life beyond 100 years is to rejuvenate dead cells and restore them to functionality. A cell does not cease to replicate because it is dead, rather a cell stay dormant and disconnected from its channel in the network inside the body. A cell that is not communicating with the rest of the body is not going to be restored to life without the proper approach to revitalize that ailing cell. A dormant cell can be restored to life as it is made up of atoms and atoms are indestructible. We have forgotten that all life in the universe is made up of matter in motion; as the theory

of relativity explains: $E=MC^2$ or said otherwise (Life) = Matter in Motion.

The answer to prolonging life lies beneath the brain and the nervous system. How important is the electrical circuitry of the brain and the body and how crucial is an understanding of neural energy as it flows throughout the human body. Cancer, Alzheimer, Parkinson's and other terminal diseases can only be cured, when cells in the affected area are electrically energized. Neural energy should flow to the nucleus of each ailing cell, for the DNA in that cell to do its work.

Research in Bio-physic of the human cell will revolutionize the future of medicine; this is the answer to healing all the mysterious diseases that we know too little about. This new approach rely solely on electrical functions of the brain as it is always communicating with the entire body through electrical signals flowing from point A to point B; the day an organ is cut off from that harmonious and synchronized network, the organ will cease to function and stays isolate without fulfilling its vital functions inside the human body.

Controlling the flow of neural energy into the areas affected by any disease, we can cure the ailment. The patient assisted by a neuroscientist who will map his brain and help stimulate diseased areas in the brain by gentle electrical vibration, to target the illness; or by using medication with controlled narcotic effect that

will stimulate affected area, to relax and allow neural energy flow to cells' nucleuses for their DNA to function.

Another important point to ponder: The brain only communicates with a cell whose DNA structure is normal. A cell whose DNA is in disarray will send a different message and will not be ionized the same way ordinary health cells do. Remember for any activity by the human, the brain always send a signal to the concerned parties (cells to be used for a function) and then the sensual activities take effect. We always evaluate an action before we perform it – we refer to the blue print of physical body that is stored in our brain. Not one single action can be taken without the knowledge that it can be accomplished successfully.

It is possible to heal oneself completely from any terminal disease by harnessing the new technique to use ones own brain energy. To the skeptics this may seems absurd to handle each individual cell, when there are over 100 trillion cells in the human body; but remember energy flow at the speed of light and we can target vast areas of the body in tenth of seconds. The speed of light is 186,000 mile/second; that is lightening fast.

Tai Chi's technique of focusing neural energy on ailing organ is a big contributor to this discipline. Psychoanalysis had also contributed to my understanding of using brain energy to heal oneself; especially the emphasis on exploring the subconscious part of our brain; the dormant thoughts and archaic mental

drives. Once we dig in our memory, we are accessing cells that have been dormant sometimes for years; once accessed, these memory areas are energized again and the individual gain awareness of his ailing thoughts and become functional.

Lastly, electronic thought me the concept of how atoms are energized and the flow of electrons and protons in a circuit. The human cell is likened to an electronic component, such as a relay, a capacitor or even a CPU (Central Processing Unit); it stores energy, which it uses to fulfill its functions.

This approach to curing Terminal Diseases, rely solely on the use of neural energy, which is scientifically asserted to exist in our brain. The energy flows throughout the entire body and it makes healing possible. Resonance or magnetism from healthy neighboring cells allows heat to sink through to these dead cells. Reaching the nucleus of dormant cells, it animates them to life; this is the only way to bring a change to the condition of DWC. The difference between a dead man and a living one is heat; there is no heat in a dead corpse.

The technique deals with the working of the cell that is the building block of human life. To make this approach a reality, we need to focus our attention on the working parts of a human cell and gain an understanding of how DNA is affecting our cell's health. The cells are chained together to form channels inside the body, forming a network; dead cells are disconnected from this

network and become tissue that is deprived of neural energy and blood circulation. This isolation creates clusters of dead cells, which I term: Dead Weight Cells (DWC) or dead flesh.

Healing through the Brain is a reality and we have to test it and see the transformation that could happen to our lives; we will be healed from the state of disease and pain to a condition of health and wellness. As human beings we have the ability to heal ourselves.

In the first few chapters of this book, I am presenting what today's research explains about DNA and its importance in enhancing human life; in the last chapters, I will be proposing the approach to debugging DNA inside the human and making it useful to improve life. If you are familiar with the DNA process in biological labs, you can skip to chapters fourteen and what follows where I explain this system of healing. In chapter sixteen I am putting forward the new system of Debugging Human DNA. I explain the phenomena of DNA and underline the importance of electrical functions in the human; they are the way by which we can restore ourselves to health and guaranty a longer life.

Submission Date: 9/27/2017

Name: David Hunter

Name of Organization: Harvard University

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Benefits high. Risks low. The requirement that people accessing the data should not attempt individual identification should be a condition of access.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Reasonable arguments that certain data may compromise privacy or stigmatize a group should be listened to.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

Reasonable arguments that certain data may compromise privacy or stigmatize a group should be listened to.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

An ongoing issue however, occurs when NIH-supported data are jointly analyzed with data from other countries not supported by NIH funding. Colleagues in these other countries report (a) they are told that they are not under data sharing obligations (b) even when their funding agencies have such obligations on paper they do not enforce them and (c) their academic institutions are anxious about public data sharing and forbid it. It would be very helpful if NIH would consider joint policies with the EU, UK MRC, CRUK, Wellcome Trust etc. to reach uniform policies. Otherwise the summary results have to be redacted to NIH-funded data only before posting.

Submission Date: 9/29/2017

Name: Bruce M Psaty

Name of Organization: Univ of Washington

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

The risks are probably minimal although this approach violates some existing consents to preserve privacy.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

The rapid access tier does indeed provide more rapid access. But what happens in the event of a breach? For dbGaP approvals, the institution was signing off, and any problems could be brought to the institution where the offending investigator works. What are the recourses for a breach of the agreement under this mechanism?

Submission Date: 10/4/2017

Name: NHGRI-EBI GWAS Catalog

Name of Organization: EMBL-EBI

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

The NHGRI-EBI GWAS Catalog recognises the benefit of access to allow follow on studies that analyse summary statistics and support this proposal. These access procedures enable maximal mileage for science funding as the results can be shared broadly for further opportunity and usage. Ultimately it is of significant benefit to scientific research and the community as a whole if these data are shared.

In addition to adhering to the FAIR data principles (data must be findable, accessible, interoperable and re-usable (<https://www.force11.org/group/fairgroup/fairprinciples>)), the importance of data standards cannot be understated. For summary statistics in particular, phenotype data, ancestry data and any relevant study metadata are imperative for the purposes of large scale data integration.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

The risk of sensitive data being held under controlled access, such as in resources like dbGaP or EGA is very low usage. We estimate that very few controlled access studies are linked to public summary statistics, thus the data are not easily Findable for community use. The GWAS Catalog would like to link study level metadata with public summary statistics. This would complete data held by secure resources such as dbGAP and EGA.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

The NHGRI-EBI GWAS Catalog fully supports sharing of genomic summary results. We already share some summary statistics for a subset of our data (1.5% of studies) and we

are expanding this effort. This has been requested by the Catalog users, data depositors and the wider community, represented by our advisory board. We are therefore contacting authors to ensure we promote access to these data.

We would stress the limitations of multiple study-specific strategies for data sharing. The benefits of sharing the data are hampered if the user must find the data in a fragmented way, if there are disparate Terms of use, if there is a lack of clear licences for data. These limit the 'FAIR'ness of the data and present challenges for the user community, particularly those in industry.

Submission Date: 10/4/2017

Name: Federation of American Societies for Experimental Biology (FASEB)

Name of Organization: Federation of American Societies for Experimental Biology (FASEB)

Type of Organization: Professional Org/Association

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):



October 4, 2017

Office of Science Policy
National Institutes of Health
6705 Rockledge Drive, Suite 750
Bethesda, MD 20892

Dear NIH Science Policy Team:

The Federation of American Societies for Experimental Biology (FASEB) appreciates the opportunity to comment on the National Institutes of Health's (NIH's) "Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy" ([NOT-OD-17-110](#)). FASEB is composed of 31 scientific societies representing over 130,000 biomedical and biological investigators. Earlier this year, NIH sought stakeholder input on the Database of Genotypes and Phenotypes (dbGaP), and we appreciate your consideration of these comments in the development of this policy proposal.

FASEB commends NIH's commitment to balancing the risks and benefits of expanded access to genomic summary statistics within dbGaP and other NIH genomic databases. In [comments](#) submitted on April 5, 2017 (appended), FASEB offered cautious support for increased access to these data products provided that (1) there is an educational component on appropriate and rigorous use; and (2) summary statistics from more sensitive studies remain only available through a controlled-access system. We thank NIH for addressing our concerns in this policy proposal and encourage the creation of additional informational resources on best practices for research and clinical application of these data products to further promote rigorous use.

As NIH develops the new "rapid access" interface, we recommend NIH again consult with stakeholders to ensure the platform achieves the values and benefits articulated in this policy proposal. Please do not hesitate to contact me if FASEB can provide further assistance.

Sincerely,

Thomas O. Baldwin, PhD
FASEB President

Submission Date: 10/4/2017

Name: Paula Appel

Name of Organization: student of Drexel University

Type of Organization: University

Role: Member of the Public

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I feel that if a participant does participate they are allowing NIH to use the information that is collected for scientific opportunity.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

The information being submitted and shared is being done in a controlled access with case sensitive information. I feel that the information is safe that way and it benefits science while protecting the participant.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/5/2017

Name: Emmanuel Mignot

Name of Organization: Stanford University

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I don't see any risk. Summary statistics should always be made available.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I am not sure in what cases summary statistics could be sensitive except maybe in the context of rare ethnic groups or small number of individuals.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

Good proposal

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

No comment

Submission Date: 10/10/2017

Name: Matthew Parker

Name of Organization: Sheffield Diagnostic Genetics Service

Type of Organization: Healthcare Delivery Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

click through barriers are a risk to the use of these resources

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Please don't add click-through on sites like genomAD

Submission Date: 10/10/2017

Name: Prof Cathy Abbott

Name of Organization: University of Edinburgh

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

IGMM

Western General Hospital

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

I am writing as an avid user of the ExAC and gnomAD databases from the UK, a statement which in itself testifies to the global reach of the proposed data management policy. I use ExAC and gnomAD databases regularly and have benefited enormously from the data therein and the insights it provides into human genetic disease.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of

any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Cathy Abbott

Submission Date: 10/10/2017

Name: Jason Merkin

Name of Organization: KSQ THERAPEUTICS, INC.

Type of Organization: Biotech/Pharmaceutical Company

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

I am writing as an avid user of the ExAC and gnomAD databases. [Please provide a brief description of your use of these resources and their benefits to your research]

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are

especially vulnerable to harm from possible reidentification.

Sincerely,

Jason Merkin

Submission Date: 10/10/2017

Name: Daniele Cassatella

Name of Organization: University of Lausanne

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I am writing as an avid user of the ExAC and gnomAD databases. I use it for my studies in the research of pathogenic variants in a rare disease called Kallmann syndrome. Without those database, I would have few information about the frequency of the variants I encounter in the general population.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Daniele Cassatella, PhD

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access.**

Risks and benefits may relate to participant protection issues and/or scientific opportunity

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/10/2017

Name: Ian Quigley

Name of Organization: Recursion Pharmaceuticals

Type of Organization: Biotech/Pharmaceutical Company

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Please don't implement this - click-throughs are likely to slow everyone down.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

This seems like a handy idea, especially for researchers. It may also help the public appreciate the work being done.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

As a researcher, I'd like access to these data to be paramount. As a private individual, I don't have a problem with my genome being used for good.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Thanks for being sensitive to all these issues!

Submission Date: 10/10/2017

Name: Robert Castelo

Name of Organization: Pompeu Fabra University, Barcelona, Spain

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):

Doctor Aiguader, 80
08003 Barcelona
[Tel.] +34 93 316 35 01
[Fax] +34 93 316 09 01
www.upf.edu

Barcelona, October 10th, 2017.

To whom it may concern,

I am writing as an avid user of the ExAC and gnomAD databases. They are crucial for my research on genetic disease by using them to help identifying potential pathogenicity of genomic variants.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction – there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,



Robert Castelo
Associate Professor
Dept. of Experimental and Health Sciences
Pompeu Fabra University
PRBB
Dr. Aiguader 88
08003 Barcelona
SPAIN
email: robert.castelo@upf.edu

Submission Date: 10/10/2017

Name: Frederick Roth

Name of Organization: University of Toronto

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I endorse Daniel MacArthur'd objection to click-through agreements on summary data as described here:

<https://macarthurlab.org/2017/10/10/response-to-proposal-to-update-data-management-of-genomic-summary-results-under-the-nih-genomic-data-sharing-policy/amp/>

They will have little benefit and potentially substantial harms in terms of limiting programmatic access to summary results

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/10/2017

Name: Brooke Wolford

Name of Organization: University of Michigan

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):

To whom it may concern:

I am writing as a user of the ExAC and gnomAD databases. I am frequently needing to explore allele frequencies in ExAC to compare to data sets that my lab uses to study cardiovascular traits.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction – there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Brooke Wolford

Submission Date: 10/10/2017

Name: Helen Naylor

Name of Organization: Vanderbilt University Medical Center

Type of Organization: Healthcare Delivery Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I am writing as an avid user of the ExAC and gnomAD databases. I use this resource to verify the frequency and orientation of mutations on a daily basis.

I believe that the new "Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy" is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/10/2017

Name: Jessica Chong

Name of Organization: University of Washington

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I am writing as both a researcher running MyGene2 and Geno2MP, two genomic data sharing resources, and as an end-user of genomic data sharing resources such as gnomAD and EVS. These resources are absolutely critical to rare disease researchers as the chance that any single investigator has multiple cases with the same "causal gene" decreases as we study conditions that are increasingly rare.

Requiring users of rapid-access data resources to do a click-through agreement does not make sense. Click-through agreements make programmatic access (via APIs) very difficult if not impossible to implement. This inhibits the spread of information shared through sites such as ours and makes it difficult for other groups to reuse our data. MyGene2, in particular, serves families/patients with rare diseases. They have specifically chosen to share their data publicly because they want to find others who have the same candidate genes/diseases. Requiring a click-through agreement would mean that we will be unable to fulfill one of our chief promises to these families -- that our site will help them find as wide an audience as possible in order to increase their chance of finding others with the same disease. For example, MyGene2 entries created by families would no longer be discoverable via Google.

Most importantly, click-through agreements would impose limitations on legitimate users while not preventing any harmful usage of shared genomic data. Any theoretical bad actor who is intent on malicious usage will simply ignore the click-through agreement, while legitimate users (researchers, clinicians, patients, and the public) are significantly inconvenienced/prevented from reusing the data in ways that benefit the entire community.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/10/2017 14:32

Name: Andrew McDavid

Name of Organization: University of Rochester

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I am writing as a statistician who has found easy re-use and meta-analysis of published genomic findings to be great value. I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. Click-through agreements complicate the programmatic reuse of summary statistics, but there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

Since the click-through agreements may only be providing a fig-leaf of security against a hypothetical risk of re-identification, but has real deleterious effects on the re-use of aggregated data, I urge the abandonment of this requirement in the final, amended GDS policy.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/10/2017

Name: JAMES Priest

Name of Organization: Stanford

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I dislike the click through user agreement

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/10/2017

Name: Eric Weitz

Name of Organization: Not applicable

Type of Organization: Not Applicable

Role: Member of the Public

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I am writing as a user of the ExAC and gnomAD databases. I spend a significant amount of my personal time, outside of work, developing tools for genomic data visualization and analysis that use genomic summary results from databases like ExAC and gnomAD.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Eric Weitz

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access.**

Risks and benefits may relate to participant protection issues and/or scientific opportunity

I am writing as a user of the ExAC and gnomAD databases. I spend a significant amount of my personal time, outside of work, developing tools for genomic data visualization and analysis that use genomic summary results from databases like ExAC and gnomAD.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Eric Weitz

3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive

I am writing as a user of the ExAC and gnomAD databases. I spend a significant amount of my personal time, outside of work, developing tools for genomic data visualization and analysis that use genomic summary results from databases like ExAC and gnomAD.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to

participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Eric Weitz

4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies

I am writing as a user of the ExAC and gnomAD databases. I spend a significant amount of my personal time, outside of work, developing tools for genomic data visualization and analysis that use genomic summary results from databases like ExAC and gnomAD.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Eric Weitz

Submission Date: 10/11/2017

Name: Keren Carss

Name of Organization: University of Cambridge

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I am writing as a regular user of the ExAC and gnomAD databases. My primary use is to use the frequencies to help interpret genetic variation in patients in my cohort with rare diseases.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely, Keren Carss

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/11/2017

Name: Kevin Stachelek

Name of Organization: University of Southern California

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I support the release of aggregated genomic data to further biomedical research and enable training of biomedical researchers in computational techniques relevant to breakthrough scientific work. I oppose unnecessary click through agreements which would discourage data sharing.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I support the release of aggregated genomic data to further biomedical research and enable training of biomedical researchers in computational techniques relevant to breakthrough scientific work. I oppose unnecessary click through agreements which would discourage data sharing.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

I support the release of aggregated genomic data to further biomedical research and enable training of biomedical researchers in computational techniques relevant to breakthrough scientific work. I oppose unnecessary click through agreements which would discourage data sharing.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

I support the release of aggregated genomic data to further biomedical research and enable training of biomedical researchers in computational techniques relevant to breakthrough scientific work. I oppose unnecessary click through agreements which would discourage data sharing.

Submission Date: 10/11/2017

Name: David von Schack

Name of Organization: Pfizer Research and Development

Type of Organization: Biotech/Pharmaceutical Company

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

I am writing as an avid user of the ExAC and gnomAD databases. I am a scientist in the Pharmaceutical Industry who regularly uses databases such as ExAC or gnomAD for new therapeutic target identification and validation purposes. Summary statistics from large cohorts across multiple ethnicities provides a unique opportunity to understand genetic selection of certain variants in genes/pathways of interest for drug target prioritization purposes.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of

any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Submission Date: 10/12/2017

Name: Mette Nyegaard

Name of Organization: Aarhus University

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I am writing as an avid user of the ExAC and gnomAD databases. In my research I am using the ExAC and gnomAD database every day and they have become the most central tool I have for variant interpretation. I do both clinical sequencing (searching for strong pathogenic variants) as well as looking for the molecular mechanism underlying a range of complex diseases, such as coronary artery disease and psychiatric disorders.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Mette Nyegaard

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/13/2017

Name: Amanda Spurdle

Name of Organization: QIMR Berghofer MRI

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity.**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):

Office of Science Policy (OSP), National Institutes of Health, 6705 Rockledge Drive, Suite 750,
Bethesda, MD 20892

13 October 2017

To whom it may concern,

I am writing as an avid user of the ExAC and gnomAD databases. I work in the field of variant classification of cancer (as head of the ENIGMA consortium, which is a key collaborator in the GA4GH BRCA Challenge project), and now other disease genes (as part of the Australian Genomic Health Alliance. Ready access to large control datasets is critical to my clinically-directed research.

I believe that the new "Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy" is a step in the right direction – there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.



Amanda Spurdle, Ph.D.

Associate Professor Amanda Spurdle
Group Leader, Molecular Cancer Epidemiology
QIMR Berghofer Medical Research Institute
Amanda.Spurdle@qimrberghofer.edu.au

Submission Date: 10/16/2017

Name: Lori

Name of Organization: Not Applicable

Type of Organization: Not Applicable

Role: Member of the Public

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

NIH-funded research information should be available only to Americans, since we have paid for it through our taxes. Although I don't understand the exact information which would become publicly available, I don't think it should have broad access for several reasons.

1. Who is going to enforce the click through agreement, and how? 2. What will prevent rogue nations and individuals from using this information to create biological weapons and/or other types of targeted weapons/diseases based on genomic traits and susceptibilities? 3. Can it really be safeguarded from hacking?

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

See above, but who is to say that the submitting institution has the best handle on what is truly sensitive information and sensitive to who? Insurance companies gaining access to information that may not be considered sensitive at the moment may result in future instances of individuals and their descendants being denied coverage based on genetic information analysis performed by or for insurance companies. It is a wild west right now without adequate regulation.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

See #2.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

The Equifax breach has exposed most of our personal information. What happens when our genomic information is also hacked because it was "open" on the internet? Because of EMA Policy 70 and the future Clinical Trials Regulation, pharmaceutical firms are spending tons of money to de-identify and anonymize patient data that supports marketing approval applications for drugs. Who is going to pay to do this for NIH-funded studies and data - or make sure that the re-identifying "keys" and other information is sufficiently protected if multiple breaches across databases occurs?

Submission Date: 10/16/2017

Name: Michael Hoffman

Name of Organization: University of Toronto

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I strongly support NIH's effort to make genomic summary statistics more widely available. This is an unqualified good for lifesaving scientific research.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I strongly oppose the proposal to lock summary statistics behind a click-through agreement. This will slow projects that might use these data, costing money and delaying scientific research. A balanced risk management analysis should show that (1) there is little risk to misuse of summary statistics, and (2) a click-through agreement provides zero meaningful reduction to risk.

A click-through agreement provides zero protection against nefarious actors. It only provides costly "security theater" and introduces unnecessary barriers to the full utilization of these datasets.

I also oppose an alternative suggestion to require use of a secret API key obtained through a click-through agreement. In my experience as a researcher and educator, getting these keys to work is usually one of the most difficult parts of learning how to use an authenticated API.

In summary, requiring a click-through agreement costs much and accomplishes little.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/16/2017

Name: Alicia Martin

Name of Organization: Broad Institute & Massachusetts General Hospital

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):



Harvard
Medical
School



Massachusetts
General Hospital

Alicia R. Martin, PhD
Postdoctoral Research Fellow

Department of Medicine
Analytic and Translational Genetics Unit
Massachusetts General Hospital
Simches Research Center
185 Cambridge Street
CPZN-6818
Boston, MA 02114

Phone:
Fax: 617-726-5937
Email:

October 16, 2017

To Whom It May Concern:

I am writing to express my strong disapproval of the NIH Genomic Data Sharing Policy's Access Model for Genomic Summary Results. This proposal fundamentally locks summary statistics from genetic data behind "click through" agreements. The leading efforts of large-scale consortia currently make summary statistics openly available concomitantly with publication, which has enabled rapid progress in genetic research. Summary statistics from genetic studies are analogous to distributions that summarize any other type of biomedical data that are freely shared in the biomedical community. They have the potential to inform the spectrum of variation in healthy and disease cohorts, and this information needs to be easily accessible to enable rapid research development towards translational therapies. Genetic summary statistics are not special compared to other biomedical summaries, and should not be put behind walls that hinder research progress.

Simply put, summary statistics do not contain information that can be detrimental to a single individual. The benefits of immediate release without jumping through bureaucratic hoops strongly outweigh the minimal risks. Moreover, they have the direct potential to advance precision medicine efforts to provide information about risk of a given disease. The proposed "click through" agreements will not hinder nefarious users, but will hinder progress of genomic research.

Lastly, as a researcher who works directly with underprivileged and indigenous populations, I am enormously concerned that summary statistics from these populations would be placed behind even tighter control. Major health disparities are likely to be exacerbated where they already exist in these populations by this policy.

In summary, requiring a click-through agreement costs much and accomplishes little.

Sincerely,

Alicia Martin, PhD

Submission Date: 10/16/2017

Name: Christoffer Flensburg

Name of Organization: Walter and Eliza Hall Institute

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I was made aware of this proposal through Danila MacArthurs post here:
<https://macarthurlab.org/2017/10/10/response-to-proposal-to-update-data-management-of-genomic-summary-results-under-the-nih-genomic-data-sharing-policy/>

I am a developer of an open source analysis software for cancer sequencing data:
<https://github.com/ChristofferFlensburg/superFreq>. While I am based in Australia, my software is used worldwide, including the US. I use ExAC data as an important piece of the software, and a click-through agreement would severely complicate and slow down the process for both me as developer and for the users. As the ExAC data I re-share are not linked to phenotypes, the commonly quoted dangers of sharing genomic data do not apply. So in my case, a click-through agreement would limit cancer research in the US and worldwide, for no meaningful added protection of the participants.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction “ there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open

access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

All the best,

Christoffer Flensburg

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/17/2017

Name: Karen Yu

Name of Organization: BioMarin Pharmaceuticals Inc

Type of Organization: Biotech/Pharmaceutical Company

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Open access to summary data would benefit all communities

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Summary data should all be open access

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

Summary data including phenotypes should be an integral component of summary genomics data.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/17/2017

Name: Centre of Genomics and Policy

Name of Organization: McGill University

Type of Organization: University

Role: Bioethicist/Social Science Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

We are very supportive of the NIH proposal to expand access to genomic summary results from most studies in NIH-designated data repositories. We strongly agree with the proposal to provide protections that are proportional to the risks to participants, notably, the risks of harm to participants if they were associated with a particular trait or other information that is studied. We actually feel most of these data may even be suited to completely open access databases. However, we recognize the difficulty in drawing clear lines between levels of uncertain risk, as well as the disadvantages of separating these data into more tiers than necessary.

For the rapid access model, a browse-through agreement, rather than a click-through agreement, might be better suited to data that are intended for broad use and integration with other similar datasets. However, we understand that the protections provided through rapid access will need to be evaluated over time.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

We feel that perhaps the summary genomic data designated as including sensitive information could be sufficiently protected through new (less restrictive) data access models, such as the Registered Access model currently being developed by the Global Alliance for Genomics and Health (GA4GH; <http://genomicsandhealth.org>). Registered Access is an intermediate data access model that is based on broad access to multiple datasets for users within defined roles (e.g., researcher, clinical care professional), who agree to a set of standard conditions of data access and use (SOM Dyke, E Kirby, M Shabani, A Thorogood, K Kato, BM Knoppers. (2016) Registered Access: A “Triple-A” Approach. *European Journal of Human Genetics*, 24 (12): 1676-80). However, this would potentially depend on the level of risk encountered with these sensitive datasets.

3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive

We believe this is a reasonable approach. However, it might be important to maintain a level of oversight, at least until comprehensive guidance on assessing levels of risk based on the sensitivity of traits and vulnerability of groups were to be available. In particular, institutions submitting data designated as sensitive could be asked to provide a clear justification for this decision.

4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies

Submission Date: 10/17/2017

Name: Avinash Ramu

Name of Organization: WUSTL

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):

I am writing as a grad student in Human Genetics who uses resources like ExAC and gnomAD on a daily basis.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction – there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible re-identification.

Sincerely,

Avinash Ramu

Submission Date: 10/17/2017

Name: John Thompson

Name of Organization: Claritas Genomics

Type of Organization: Biotech/Pharmaceutical Company

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

When there are security/privacy rights that are at risk, it is important to institute appropriate controls to avoid harm. Providing summary data unlinked to phenotype and identifying information poses no risk to anyone. Easy access to such data is critical for helping to understand genetic disease. There should be no click-through or any other security on data that poses no risk. Putting any level of protection on the data is unwarranted and can only slow progress in helping patients and providers understand their genetic issues.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

In general, genomic summary results can be made non-sensitive and the onus should be on the submitting institution to prove that the data cannot be sanitized. If NIH or other public organization is funding the work, it should be shared as quickly and as completely as possible.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Genomic data needs to be shared for maximal value. Anything that impedes sharing needs to be clearly justified and that justification needs to be carefully considered by funders to ensure the reasons are valid.

Submission Date: 10/17/2017

Name: Ian Dunham

Name of Organization: European Bioinformatics Institute

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I strongly support the NIH's drive to promote rapid access to genomic summary results. This will promote wide re-use of the data for analyses that were not necessarily envisaged when the data was collected. In my view potential harms from this data to individuals through identification or stigmatisation are largely hypothetical and unlikely to be realised through these data. However the potential loss of opportunity by not sharing the data is genuine and significant. For instance each single mutation in a gene in a rare Mendelian disease is a valuable data point that can reveal information on the potential efficacy or safety of future drugs that antagonise the gene product in diseases not limited to the observation of origin. The benefit to the collective research enterprise and drug development of having these data available extends way beyond the individual to society-wide benefits.

I would strongly advise that, after a thorough risk assessment of the possible risks of release, that the data should be available with minimal barriers. Even a click through mechanism is a significant obstacle to programmatic access and should be avoided since it provides no real protection (if protection is needed). If we believe that the risks are minimal and out-weighed by the global benefits, we should back this by impediment free access.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies

I strongly support the motives behind this advance. There is a worrying trend in the field moving away from the open data ethos of the genome projects, ENCODE and 1000 genomes etc. to creation of data fiefdoms with barriers to entry supported by over blown concerns of data protection for DNA sequence data. All possible effort should be focussed on enabling the global research effort to discover new mechanisms and medicines by capitalising on these valuable resources.

Submission Date: 10/17/2017

Name: J.A. Lathrop

Name of Organization: N/A

Type of Organization:

Role:

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Privacy must be protected but if Total numbers from a pool of genetic test and no names of individuals is submitted to NIH there should be no problem.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

NIH should not be given names. If they would like to study individuals the testing organization should get permission from the individual FIRST only then should NIH be able to contact them.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

All datasets should be designated as sensitive unless tested parties agree that it is not sensitive.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Personally I probably would not care if my name and data were shared without restriction for research purposes. My family history is such that I would love for more data on prevention and cures to be available. There are other members of my family that simply don't share that trust and in their cases that should be respected.

Submission Date: 10/17/2017

Name: Clare Turnbull

Name of Organization: Institute of Cancer Research

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Risks of click through access to ExAC/gnomAD: thwarting integration of ExAC/gnomAD variant counts into other useful databases. Misclassification of individual variants leading to harm for my patients.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Risks of restricted access to ExAC/gnomAD: thwarting research progress. Misclassification of individual variants leading to harm for my patients.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

Only individual level large scale sequencing data (ie potentially identifiable) should be treated as sensitive. Summary level data eg ExAC/gnomAD is not in the slightest bit sensitive.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

I am writing as an avid user of the ExAC and gnomAD databases. We are using these databases to populate our centrally integrated UK database on cancer susceptibility genes (CaVaDa) against which we have integrated a wealth of disparate variant-level datasets (<http://cavada.dynalias.org/cavada/>). Our database is used by many clinicians and researchers in the UK to access variant level data to inform clinical variant interpretation. ExAC/gnomAD variant frequencies form the backbone of information for each variant.: we need to be able to download these data in their entirety to populate our database and inform our analyses.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Clare Turnbull MD PhD MA MSc MRCP MFPH DCH

Professor of Genomic Medicine | Clinical Lead for Cancer Genomics for 100,000 Genomes Project | Honorary Consultant in Clinical Cancer Genetics

Genomics England | Queen Mary University of London | Dawson Hall, London, EC1M 6BQ | T: 0207 882 6393/0207 882 6392 | Institute of Cancer Research| SM2 5NG | T: 020 8722 4487 | Guys and St Thomas' NHS Trust | SE1 9RT.

Submission ID Date: 10/17/2017

Name: Obi Griffith

Name of Organization: McDonnell Genome Institute, Washington University

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

I am writing as an avid user of the ExAC and gnomAD databases. My group regularly uses these datasets to distinguish important cancer variants from common germline polymorphisms.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with

the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Obi Griffith

Submission ID Date: 10/17/2017

Name: Gopala Tumuluri

Name of Organization: UPMC

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I am writing as an avid user of the ExAC and gnomAD databases. Our organization uses these datasets as a reference in analyzing research study data produced within our system through gene sequencing.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Gopala Tumuluri

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/17/2017

Name: Dana Bis

Name of Organization: University of Miami

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I am writing as an avid user of the ExAC and gnomAD databases. I pursue gene discovery in inherited peripheral neuropathies at the University of Miami. When determining the pathogenicity of a rare variant, access to a large public database from unaffected individuals is absolutely necessary. The primary filtering step when analyzing a patient's DNA sequencing data is to remove variants that exist above a threshold frequency within these databases, such as GnomAD.

I believe that the new "Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy" is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Dana Bis

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/18/2017

Name: Richard Bagnall

Name of Organization: Centenary Institute

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):

I am writing as an avid user of the ExAC and gnomAD databases. I use these resources to identify clinically relevant DNA variants in patients with inherited heart diseases.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction – there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Dr. Richard D. Bagnall

Submission Date: 10/18/2017

Name: Prof Chris Semsarian

Name of Organization: University of Sydney

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):

I am writing as an avid user of the ExAC and gnomAD databases. I use these resources to identify clinically relevant DNA variants in patients with inherited heart diseases.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction – there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Professor Chris Semsarian
University of Sydney, Australia

Submission Date: 10/18/2017

Name: Hugh French

Name of Organization: Sydney Medical School

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):

I am writing as an avid user of the ExAC and gnomAD databases. These resources and the access of them are essential for my research into Hypertrophic Cardiomyopathy.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction – there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Hugh John French, PhD.

Submission Date: 10/18/2017

Name: Kodama Yuichi

Name of Organization: DDBJ Center, National Institute of Genetics

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

To provide more rapid and easy access to the summary results will benefit whole scientific community. Researchers interested in only summary results will be able to obtain and analyze the summary results without consuming lots of time for the same process required for accessing the dbGaP individual-level data. In addition, the new rapid access model will advance meta-analysis studies by combining the summary results from multiple studies. Risks are considered to be little. To identify if specific person is included in the summary results or not, genome sequencing or genotyping data of that particular person is necessary. It would be extremely rare, someone already knowing the genotype of target person, attempt to analyze whether the person is included in the public dbGaP summary result datasets. For safe guards, the proposed simple one-click registration would be enough. In addition, it is important to provide options to dbGaP submitters to hide their summary results from indigenous people, phenotypes possibly linked to stigma.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

There should be the options to remain the summary results in controlled-access to respect and protect participant's informed consents. This could be unfavorable for science advancement, but scientific community should accept this if participant want to put their summary results into controlled-access.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

Sensitivity of data is complex and depends on the study context. It could be difficult to define sensitive information uniformly, so this needs to be defined by submitters. Under

the NIH GDS policy, it is appropriate to allow submitters to define sensitive information in their submitting data. But NIH needs to check the appropriateness of sensitive information categorization by submitters.

4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies

To provide the summary results in GRU (General research purpose) will further advance research. The summary results are considered to be more open than underlying individual-level data, so provide the summary results in less stringent data use restrictions, for example, individual-level data for cancer research only but the summary results derive from the data is GRU, will further promote science.

Submission Date: 10/18/2017

Name: Natalie Nowak

Name of Organization: Centenary Institute

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):

I am writing as an avid user of the ExAC and gnomAD databases. I use these resources to identify clinically relevant DNA variants in patients with inherited heart diseases.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction – there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Natalie Nowak

Submission Date: 10/18/2017

Name: Emma Singer

Name of Organization: Centenary Institute of Cancer Medicine and Cell Biology

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

N/A

Additional Comment (attachment):

I am writing as an avid user of the ExAC and gnomAD databases. I use these resources to identify clinically relevant DNA variants in patients with inherited heart diseases.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction – there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Emma Singer

Submission Date: 10/18/2017

Name: Laura Yeates

Name of Organization: Centenary Institute

Type of Organization: Nonprofit Research Organization

Role: Medical Provider

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):

I am writing as an avid user of the ExAC and gnomAD databases. I use these resources to identify clinically relevant DNA variants in patients with inherited heart diseases.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction – there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Laura Yeates
Genetic Counsellor
Sydney, Australia

Submission Date: 10/18/2017

Name: Jodie Ingles

Name of Organization: University of Sydney

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):

I am writing as an avid user of the ExAC and gnomAD databases. I use these resources to identify clinically relevant DNA variants in patients with inherited heart diseases.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction – there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Dr Jodie Ingles
Cardiac Genetic Counsellor and Researcher
University of Sydney

Submission Date: 10/19/2017

Name: Michael Boehnke

Name of Organization: University of Michigan Department of Biostatistics

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

For common diseases and large populations, there are no meaningful risks for sharing genetic summary results. These results are of high value to the scientific community, and so making them widely available would be a great value to the scientific community and to society which will benefit from facilitated scientific discoveries. Summary results are much more valuable when shared openly on the web than through a click-through mechanism because they will be easier to obtain and more widely used. I have attended two NIH meetings on this topic, one that I co-chaired, and both came out strongly in favor of fully-open access with no click through needed.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

I could imagine this being useful for rare diseases, stigmatizing conditions, or special populations. For this reason, I would allow such an exclusion to open access. I would also monitor how frequently and under what circumstances this exclusion is employed to seek to ensure it is not overused.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

For new studies, NIH should encourage investigators to choose broad consent as the default unless there is a strong reason why it is not appropriate.

Submission Date: 10/19/2017

Name: Lea Davis

Name of Organization: Vanderbilt University Medical Center

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I am a strong proponent of broad sharing of genomic summary statistics. Several biological and evolutionary insights in at least five of my recent publications have been based on the use of summary statistics. I firmly believe that these insights would not have been possible without access to summary statistics. I think that the benefits of data sharing (i.e., biological discovery and increase of knowledge) far outweigh the extremely small risk of loss of confidentiality inherent in summary statistics.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I think maintaining the controlled access of sensitive or highly stigmatizing traits is reasonable. While the opportunity for benefit may be impacted in terms of subsequent scientific discovery, the participation in such studies is likely to be higher if additional protections can be offered. I think this would balance the benefit loss and ultimately balance the risk/benefit ratio.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

I think this is a reasonable approach though I am concerned about the possibility of the abuse of "sensitive" status to restrict data sharing. I think it would be helpful for the NIH to release some guidelines on this.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

I am grateful for the opportunity to comment on the new policy and am very much in favor of rapid access to summary statistic results.

Submission Date: 10/19/2017

Name: Daniel E. Weeks

Name of Organization: University of Pittsburgh

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I am writing as an avid user of the ExAC and gnomAD databases. [Please provide a brief description of your use of these resources and their benefits to your research]

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Daniel E. Weeks

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access.**

Risks and benefits may relate to participant protection issues and/or scientific opportunity

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/19/2017

Name: Amelie Stein

Name of Organization: University of Copenhagen

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I am writing as an avid user of the ExAC and gnomAD databases, which provide an invaluable background set in our comparison of pathogenic and near-neutral variants of proteins, assessment of their functional consequences and development of better predictors for clinical application.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Amelie Stein, PhD

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/19/2017

Name: Chun-yu Liu

Name of Organization: University of Illinois at Chicago

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Rapid access will promote collaborative research, fast development of science, hopefully earlier discoveries. Summary statistics impose minimum risk to subjects.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Most so-called sensitive information may not be a concern at the summary statistics level. But it should be reviewed in a case by case manner.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

It is important to have that

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Different level access could be applied to different types of data. Ideally controlled access should only be applicable to a limit types of cohorts and phenotypes.

Submission Date: 10/19/2017

Name: Bingshan Li

Name of Organization: Vanderbilt University

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I do not see major risks but the benefits of the broader sharing are HUGE!!! By the additional summary data researchers are able to perform more in-depth analyses that are not otherwise possible.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

This may depend on what sensitive information to include. I do not feel any major risks if the information is consented by the patients.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

I think it is appropriate.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Given the complicated nature of controlled access, I greatly favor summary statistics. To maximize the power of summary statistics, I am in great favor of broader sharing so that the research community is able to perform most of the analyses without going through dbGaP. If possible, more details summary results may be shared, say results for sub-phenotypes or by gender or other phenotypes available, principal components, global and local ancestry, etc., so that people may do more analyses. Phase information is currently not available but can be useful, if there is a good way to represent that across the genome. In general, I believe the research community can benefit much more if broader (even broader than what has been proposed) sharing is executed.

Submission Date: 10/19/2017

Name: Prescott Woodruff

Name of Organization: UCSF

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I believe the benefits of broadly sharing genomic summary data significantly outweigh the risks and that this is an important initiative. Current rules do not provide enough guidance and err on the side of restricting data sharing. If no actual sequence data but rather summary data is what is being shared then the main risk (that the data can not be deidentified) is significantly mitigated. I support progress in loosening restrictions on this type of summary data.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/19/2017

Name: Ben Neale

Name of Organization: Broad Institute

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

See attached

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

See attached

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

See attached

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

See attached

Additional Comment (attachment):

Benjamin M. Neale, Ph.D.
Assistant in Genetics
Analytic and Translational Genetics Unit
Massachusetts General Hospital

Assistant Professor
Harvard Medical School

Director of Population Genetics
Stanley Center for Psychiatric Research
Program in Medical and Population Genetics
Broad Institute of Harvard and MIT

Department of Medicine
Analytic and Translational Genetics Unit
Massachusetts General Hospital
Simches Research Center
185 Cambridge Street
CPZN-6818
Boston, MA 02114

Office: 617-643-5148
Fax: 617-643-3293
Email: neale@atgu.mgh.harvard.edu

October 19th,

To whom it may concern,

Thank you for the opportunity to comment on the proposed update to the access procedures for genomic summary results under the Genomic Data Sharing (GDS) Policy. We would like to commend the NIH's thoughtful consideration of how to maximize the public benefit of genomic data through open access to genomic summary results. Though concerns have been raised in the past – as pointed out in the Request for Comment (RFC) - about the identifiability of genomic summary results, these concerns have been shown to be largely theoretical¹ and importantly, we and NIH are aware of no known instances in which a study participant has been harmed as a result of open access to genomic summary results of a study in which they have participated. We are therefore very excited at the potential increased scientific value that will be enabled by the broader access to genomic summary results permitted by this revised policy.

We are, however, concerned about two particular aspects of the proposed policy change, specifically the requirement for a click-through agreement in order to gain access to the summary statistics, and the designation of some genomic summary results as “sensitive.”

Though a click through agreement may seem like a low barrier to access to genomic summary results, Dr. Daniel MacArthur has explained very clearly in his comment letter the less immediately obvious barriers it creates, particularly to application programming interface (API) driven automatic queries and programmatic access as well as the complication it creates to data re-use. Given the barriers to the very research this policy seeks to enable that a click through agreement creates, without doing anything substantive to prevent the kinds of malicious data misuse it seeks to prevent, we strongly endorse Dr. MacArthur's suggestion to clearly communicate to users about responsible use of genomic summary results, but without the requirement for a click through agreement.

In addition to the concerns raised by Dr. MacArthur, we are also very concerned about the policy's designation of some genomic summary results as “sensitive.” By creating a two-tier system, we run the risk of both reinforcing stigma around certain conditions and slowing down scientific conditions. Mental illness is an example of a highly stigmatized disease that has benefitted from the research enabled by open access to genomic summary results through consortia like the Psychiatric Genomic Consortium (PGC). Currently, summary statistics from analyses of schizophrenia, bipolar disorder and autism are all freely available online for use (<http://www.med.unc.edu/pgc/results-and-downloads>). We are very concerned that such research could be held back in the future if genomic summary results from psychiatric studies were held only behind controlled access barriers.

Thank you again for the opportunity to comment on this exciting proposed change to the GDS Policy. With a few minor changes, it will be a great boon to the scientific community.

Sincerely,



¹ Daly, Mark J. (April 23, 2017). **Is Reidentifiability a Risk?: Open versus restricted access for summary statistics.** Retrieved from: <https://medium.com/@mjdaly/is-reidentifiability-a-risk-ae62a691a7cc>

Submission Date: 10/19/2017

Name: Brad Chapman

Name of Organization: Harvard T.H. Chan School of Public Health

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Our group at Harvard Chan School helps develop an open source community framework called bcbio (<http://bcb.io/>). bcbio aggregates software tools, biological data, workflow definitions and validation sets into a unified framework that runs across multiple platforms. The value this provides is automating the process of setting up and running analysis methods. We have developed a large community built around maintaining and improving this automation, and it helps serve both large academic and industry research efforts as well as individual researchers.

The key component of this automation is direct access to associated tools and data, allowing us to develop a framework that automatically aggregates them. We learned about the proposed click through agreement through Daniel MacArthur's discussion regarding ExAC and gnomAD. These are two datasets that we use for automated variant annotation using other open source frameworks like GEMINI (<http://gemini.readthedocs.io/en/latest/>) and vcfanno (<https://github.com/brentp/vcfanno>).

Not being able to directly download and re-package/re-distribute widely useful datasets would prevent the type of automation we provide. This is a major disservice for researchers, who then need to manually download and obtain datasets of interest. As a practical example, we dealt with this for the past 3 years with Broad's GATK tools. These were available through a click through, requiring users to manually download and install them unlike the 100s of other tools we also provide. As a result we had to build special tooling for inputting the manually downloaded files and dealt with numerous issues related to problems and barriers that researchers had doing this.

A click-through creates a lot of extra work and barriers to actually performing science, rather than dealing with technical issues. For the 99.9% of users who want to make use of this public data, it enables them to work together as a community to do science faster. For the minority of users aiming to inappropriately use the data, the click through provides little to no barrier. As a result it ends up inconveniencing many users without a practical benefit. I hope you will reconsider a click through and instead provide the data in a freely accessible, redistributable manner so the scientific community can continue to find new

and innovative ways to repurpose this incredibly useful data.

Thank you for all your time and work on data access,

Brad Chapman

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/19/2017

Name: Ivan Limongelli

Name of Organization: enGenome

Type of Organization: Biotech/Pharmaceutical Company

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

I am writing as an avid user of the ExAC and gnomAD databases for research purposes.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open

access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Ivan Limongelli

Submission Date: 10/19/2017

Name: Ryan Herringa

Name of Organization: University of Wisconsin Madison

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I think this would go a long way towards advancing genetic and biological research to solve our very complex health problems. If we are to truly move forward, researchers need access to larger datasets and information. I think the risks of individual subject identification and stigmatization are pretty small compared to the likely benefits of increased data sharing

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/19/2017

Name: Stephen Faraone

Name of Organization: SUNY Upstate Medical University

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

There is a huge potential benefit. The early release of summary statistics would allow more researchers to conduct analyses to move their work forward and to plan future analyses. The only risk I know of is if the release of summary statistics could harm a minority population in a study that uses a focused ethnicity. Otherwise, risks seem minimal.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Same as above but the risks are accentuated.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

Not appropriate. There should be a consensus definition of what is sensitive.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

I have seen first hand the benefit of early release of summary statistics in my work in psychiatric genetics.

Submission Date: 10/19/2017

Name: Vivek Ramaswamy

Name of Organization: Mr.

Type of Organization: Not Applicable

Role: Member of the Public

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

I do not want click-through agreements to authorize genomic data sharing. Makes it harder for me to consume this information.

Submission Date: 10/19/2017

Name: Jamie L Wenke

Name of Organization: Healthcare Delivery Organization

Type of Organization: Scientific Researcher

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

I am writing as an avid user of the ExAC and gnomAD databases. [Please provide a brief description of your use of these resources and their benefits to your research]

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Jamie Wenke

Submission Date: 10/19/2017

Name: Karol Estrada

Name of Organization: Biogen

Type of Organization: Biotech/Pharmaceutical Company

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):

I am writing as head of the Statistical Genetics and Genetic Epidemiology at Biogen, and an avid user of the ExAC and gnomAD databases. These databases have been critical for understanding the effect of rare genetic variation in our drug discovery projects.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction – there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Karol Estrada, PhD
Head of Statistical Genetics and Genetic Epidemiology
Biogen.

Submission Date: 10/19/2017

Name: Rhonda Lassiter

Name of Organization: Ambry Genetics

Type of Organization: Healthcare Delivery Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

I am writing as an avid user of the ExAC and gnomAD databases. [Please provide a brief description of your use of these resources and their benefits to your research]

I believe that the new "Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy" is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Rhonda Lassiter

Submission Date: 10/19/2017

Name: Chris Balak

Name of Organization: TGen

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

I am writing as an avid user of the ExAC and gnomAD databases. We use this data daily in the center focusing on rare childhood disorders.

I believe that the new "Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy" is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Chris Balak

Submission Date: 10/19/2017

Name: Jack Kosmicki

Name of Organization: Harvard Medical School

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):

I am writing as an avid user of the ExAC and gnomAD databases. Both the ExAC and gnomAD databases have enabled my colleagues and I to discover widespread mutational recurrence in autism spectrum disorder, schizophrenia, intellectual disability, developmental delay, and congenital heart disease (PMID: 28191890). Furthermore, we use the summary statistics and allele frequencies in both ExAC and gnomAD to improve power for gene discovery in case/control and trio-based inherited association studies of neuropsychiatric diseases. Without these resources, we would require much larger sample sizes to identify novel genes and improve our understanding of the underlying biology of neuropsychiatric disease.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction – there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Jack Kosmicki

Submission Date: 10/19/2017

Name: Warren Cheung

Name of Organization: Children's Mercy Hospital

Type of Organization: Healthcare Delivery Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Broad access would encourage and spur coordinated scientific research, reduce duplication /redundant scientific output. Barriers to efficient/effective use of the data would defeat the advantages of this program or would make the implementation unwieldy. I support systems that provide a clear statement of agreement between parties, however would caution that it would be counterproductive to have any individual user see this multiple times.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Whenever possible, it would be ideal to provide access to thus form of summary data for research purposes to encourage scientific progress.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Encouraging sharing of research and unlocking data generated by public funding for public research use should be a primary goal. Results from NIH funded studies should be broadly available to all for public (i.e. NIH funders) for research purposes. NIH should encourage and facilitate and encourage data sharing among its funded projects

Submission Date: 10/19/2017

Name: Sanjay Chandriani

Name of Organization: BioMarin

Type of Organization: Biotech/Pharmaceutical Company

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Executive summary: the NIH is seeking comments on a new proposed policy on genomic data sharing. While there is much to like about the new policy, we are very concerned about the proposed requirement for a click-through agreement on all aggregate genomic resources (which would include heavily-used databases such as ExAC and gnomAD). Our draft response to the Request for Comments is below. If you agree with our concern, please consider replying to the Request for Comments yourself, using the template text at the end of this post if useful.

Draft response to Request for Comments

We would like to applaud the NIH for moving in the right direction with its new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy”. The rapid and open sharing of summary statistics from aggregate genomic data brings tremendous benefit to the scientific community, and the potential harms of such sharing are largely theoretical. Our own experience with the ExAC and gnomAD public variant frequency databases has shown that the benefits to academic, clinical and pharmaceutical scientists from sharing of aggregate data are substantial: The browsers have had over 10 million page views by over 200,000 unique

users from 166 countries in the past three years, and have been used by diagnostic laboratories in the analysis of >50,000 rare disease families. Even greater value will arise as a result of broader sharing of aggregate statistics as empowered by the new policy.

However, we are still very concerned by one aspect of the new Genomic Summary Results Data Sharing Policy - the creation of a new tier of access, rapid-access, which requires a click-through agreement to gain access to summary statistics. These concerns can be summarized as follows: (1) Click-throughs make programmatic access to data-sets challenging; (2) they greatly complicate or prevent multiple important types of re-use of the data; and (3) they are highly unlikely to deter anyone with genuine malicious intent. Overall, our position is that click-through agreements are a security fig leaf that gives the impression of extra protection, but actually do no good - and can do non-trivial harm. And we would like to emphasize that ExAC and gnomAD, along with other aggregate data sharing sites such as the Exome Variant Server, do not and never have had click-through agreements, and to the best of our knowledge no harm has ever come to participants as a result.

To explain those points in a bit more detail:

It is critical for summary statistic resources such as gnomAD that we allow access through programmatic interfaces (APIs) so that people can query them using software (e.g. pull out just a targeted slice of the data) and perform automated queries (e.g. pull out the frequency of a specific variant when a user loads a different web page about that variant). Most implementations of click-through agreements will prevent or greatly complicate any form of programmatic access. There are possible technical workarounds, but all of them result in some kind of barrier to programmatic access.

Probably the single biggest obstacle created by click-through agreements is that they prevent or substantially complicate data re-use. Right now anyone can download the complete list of frequencies from gnomAD and load it up in another website, or use it to build other useful web services (the complete ExAC data set has been downloaded thousands of times). With any kind of click-through agreement they either couldn't do that at all, or would have to incorporate the same agreement in their usage policy, which may be incompatible with their proposed usage.

Most importantly, click-through agreements do nothing to prevent the types of usage that are most likely to be harmful. It is worth noting that ExAC and gnomAD have existed on the web for almost 3 years and been accessed more than 10 million times without us being aware of a single incident that has any risk of harming participants. The vast majority of users are simply interested in using the data in their research. The theoretical bad actor who is interested in malicious usage is extremely unlikely to be dissuaded by a click-through agreement, nor does the click-through agreement offer any real after-the-fact protection if a malicious actor decides to do harm.

In summary, click-through agreements will degrade or destroy programmatic access and data reuse, without having any meaningful effect on participant safety. Any policy that

advocates for click-through agreements as a solution should spell out explicitly exactly what types of misuse the click-through will prevent, and should justify the barriers to data usage that would result.

We believe it would be a mistake to incorporate click-through agreements into any NIH-wide policy. Instead, we suggest that the NIH require clear wording about the responsible use of aggregate data (such as avoiding reidentification) on all websites sharing aggregate genetic data, perhaps with a link on every page, but with no click-through barrier. This would provide a reasonable balance between serving the needs of the research community and protecting the public trust.

Signed,

Daniel MacArthur.

A request for gnomAD users and supporters

For any member of the ExAC/gnomAD community who agrees that the public sharing of summary statistics is both harmless to participants and of great benefit to science, we urge you to read the new policy proposal here, and to respond to the NIH's Request for Comment here by October 20th.

Feel free to edit or use the text below:

I am writing as an avid user of the ExAC and gnomAD databases. [Please provide a brief description of your use of these resources and their benefits to your research]

I believe that the new "Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy" is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Sanjay

Submission Date: 10/19/2017

Name: Michael Squires

Name of Organization: Law Student, S.J. Quinney College of Law

Type of Organization: University

Role: Member of the Public

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

The procedures that govern data sharing, or access procedures for the genomic summary results under the Genomic Data Sharing (GDS) Policy must be viewed through the lens of both the goals of the National Institutes of Health (NIH), as well as the more specific, but similar goals of any genome wide association studies (GWAS). Indeed, First it is important to consider the goals of the National Institutes of Health (NIH). Some of these goals include “foster[ing] creative discoveries, innovative research strategies [â€] expand the knowledge base in medical and associated sciences in order to enhance the Nation’s economic well-being and ensure a continued high return on the public investment in research.” Second, and similar to the goals set by NIH are the goals must also be viewed in a more narrow, but relevant context of the goals of a genome-wide association study (GWAS) repository: (1) improve health; and (2) maximize public investment. The sharing of genome-based research with a broad audience will permit our nation’s scientific understanding of the human condition to grow and improve, and lead to the development of better technologies, while reducing the costs associated with the development of these technologies. However, these two ideas must be managed in the context of the applicable federal laws and regulations that govern and protect research participants’ privacy interests. For purposes of these comments, the GDS Policy applies to all large-scale data research projects funded by NIH like GWAS.

Presently, there are two tiers of access available: (1) unrestricted or open access; and (2) controlled access. From my perspective, this new proposal intends on providing a hybrid of the two - rapid access. Users will be required to agree to the terms set forth in a ‘click-through agreement.’ Concerning the degree to which access is granted, the rapid access and ‘click-through agreement’ must comport with the existing NIH framework that governs data sharing. While I share thoughts on the process concerning data submission in Question 2, the framework for how data is submitted needs to also be considered in conjunction with the degree to which access is granted, and what may need to be added to the ‘click-through agreement.’ Addressing the risks and benefits of rapid access is beneficial, but so is any action that is taken even at the very beginning when an NIH grant is awarded, e.g. What is included in the Genomic Data Sharing Plan that is

submitted to NIH? What role will an institution's IRB play in providing controls for ensuring that the genomic data that is generated is done so in accordance with the Common Rule, HIPAA, and other applicable NIH rules and industry standards.

First, NIH needs to consider reevaluating and updating their Data Use Certification (DUC) Agreements. It goes without saying that the purpose behind the rapid release of datasets, along with a shorter 'click-through agreement' is to accelerate the speed at which datasets may be requested. However, the DUCs that allow a principal investigator (PI) to request and subsequently use a genomic study dataset are designed in such a unique way so as to provide accountability and transparency so that the sensitive information that is contained in their research is protected. DUCs require that those who request access to controlled access datasets are going to use the information that they requesting responsibly. To reduce these to a 'click-through agreement' is somewhat concerning, though I do believe that it is possible to have an enhanced 'click through agreement' that would still incorporate the information already included in a DUC, but in a reduced format that could certainly be more user friendly. That being said, ensuring that the information included in DUCs is also incorporated into the 'click through agreement' is critical.

The 'click through agreement' as proposed requires that those seeking to use GWAS datasets are confirming that the user: (1) has reviewed the information resources describing the appropriate uses of genomic data, including genomic summary results; (2) will not attempt to re-identify or contact any individual or group within a study population; and (3) will promote scientific research or health through the use of the genomic summary results. Even still, these three components not begin to encompass the information that is contained in a DUC. Some recommendations that NIH may want include with the 'click through agreement' could be: (1) Providing specific details on security measures that the PI requesting the dataset will use; (2) What a PI requesting information would need to do with the information, or how access would be treated following the use of the information; (3) What oversight would be applied to PIs should they violate the DUC; and (4) How would an individual IC address the sharing of data that might contain particularly sensitive or potentially stigmatizing information on participating individuals. In summation, to what extent would the 'click through agreement' as proposed incorporate these additional requirements that already exist in DUCs? I believe that the information that a requesting institution makes in their data access request, and subsequent DUC is important to mitigate against improper data use.

Second, and in conjunction with those data management safeguards incorporated in at DUC, are the seven elements included as part of the NIH Code of Conduct for Genomic Data Use. The 'click through agreement' as presently proposed contains a couple of these elements. Namely, it requires that a PI will not attempt to identify or contact any individual in the study, but a review of, and incorporation of these seven principles reaffirm what is required as part of a DUC, and will ensure that those investigators who are requesting access to a dataset meet some requirements more than what appear to be outlined in the proposal. It will avoid the risks of PIs misusing data, but still allow for some degree of faster access to the dataset.

2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity

Generally the maintenance and subsequent protection of highly sensitive personal information has been a concern since 1974 when the Congress passed the 1974 Privacy Act. They recognized then, as subsequent laws and regulations reflect now, the high level of importance of securely storing personal information in this increasingly digital age. NIH should weigh the public good that comes as a result of genomic information versus the risk of having personal information maintained in the rapid access format. This arrangement must therefore necessarily incorporate the following considerations in the context of the Common Rule, and HIPAA: (1) IRB approval for the use of data as described in DUC and the extent to which protected health information (PHI) will be maintained so as to avoid HIPAA violations especially when a researcher is also an individual's health care provider; (2) The informed consent of individuals so that minimal risk is ensured with the new rapid access model, and 'click through agreement';

First, the Common Rule insists upon that the IRB approval process is an integral first step for any research project involving human subjects. IRBs are responsible for determining whether or not a research project can even move forward, and are required to include written procedures for the risks associated with a project that is supported by the federal project. This would relate to the collection of, but also the storage and maintenance of this information. As proposed the new rule would allow institutions that have particularly stigmatizing or sensitive personal information to submit Institutional Certificates that would include the requirements outlined by the Common Rule. Specifically, these certificates require that an institution have its IRB affirmatively approve the sharing and maintenance of the genomic data of its research, and that this approval include the way in which the data will be stored and disseminated. This presently includes the option for unrestricted or controlled access data sharing. I believe that the proposed alternative to allow institutions submitting particularly sensitive or stigmatizing information to further restrict their data from the rapid access model is appropriate. IRBs have been established as gate keepers for information, and allowing them the ability to "opt out" of allowing stigmatizing data to be shared is not only consistent with the Common Rule, but facilitate things administratively for NIH.

Second, while the NHGRI Workshop "Sharing Aggregate Genomic Data" concluded that there is minimal risk associated with loosened data access standards such as the proposed rapid access model, what is NIH capable of doing to ensure that it pose a minimal risk to participant data? Assumedly, the data that will be exchanged is de-identified which would already place the genomic information at a minimal risk level however, the proposed changes include systematically computed statistics to determine the risks associated with re-identification of information. These statistics will inform the NIH as to the risks that are associated with the storage of, and re-identification of sensitive personal data. As outlined in the proposal, these summary statistics will be available

through the repository's website and consequently public. My concern is how will NIH determine for purposes of deciding what degree of risk is proportional, and/or appropriate for the data being stored? What thresholds will be established, or will this be on a case-by-case basis? Referring to my first comment in this section, would NIH still be able to allow rapid access to information that is potentially sensitive, but of a very low risk of being re-identified? Further, a majority of the Common Rule restrictions are triggered if the level of risk for data is above a "minimal risk" level. For consistency purposes, does NIH propose to maintain the same definition for minimal risk as is included in the Common Rule definition and would NIH need to establish well-defined thresholds that clarify the risk proportionality discussed in the proposed changes?

Additionally, the Common Rule requires that a PI obtain informed consent for each human subject that meet specific requirements. While the newly finalized modifications to the Common Rule approved in 2017 make some minor changes to the way in which informed consent may be obtained, the Common Rule does require that participants be provided (1) A description of any foreseeable risks; and (2) How the confidentiality of a participants records will be maintained, amongst other requirements. To the extent that the proposal seeks to change the way in which access is granted for summary results, and that access is weighed in function to the risk associated with that data, how does the NIH anticipate modifying informed consent for subjects into the future? Also would NIH require that those institutions seeking to grant rapid access, provide information to participants about the systematically computed statistics associated with their data, and the way in which risk proportionality would be determined? One of the reasons for the revision of the Common Rule was to clarify the informed consent process, but I fear that this would not contribute to that goal. I feel that explaining the rapid access model could be confusing. Does NIH propose a way to accommodate this potential change to the way in which informed consent is obtained?

3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive

It is entirely appropriate under the new proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive. The objective of this rapid release model is to enhance the sharing of information for the advancement of public health while respecting, and protecting that information that is most susceptible to stigmatizing individuals. Providing the option for institutions to continue to specify that data can be provided only through controlled access is an important option that PIs and institutions still have under this update, and I believe that is important. NIH should also consider the following: (1) How will data sharing for sensitive data with the new rapid release model be affected if at all by the levels of data sharing; (2) Would this policy apply to both intramural and extramural research; (3) How would Certificates of Confidentiality ("COCs") be addressed to potentially still maintain confidentiality, while allowing rapid access to data.

First, NIH provides general guidance on submitting data to NIH data repositories with regard to how data that undergoes different levels of data processing. Assumedly for the purpose of the primary and secondary research, the level to which data (e.g. Level 0,

Level 1, Level 2, etc.) is processed is very important for privacy protection for those involved in human subjects, as well as the time at which the data is released. In the proposal it mentions that it is possible for an institution to specify how the data, specifically sensitive, and potentially stigmatizing PHI should be handled, and to whom should access be granted - either controlled or unrestricted access. Practically, how would this be administered through the Institutional Certification as the proposal indicates? Summary level data or secondary research would seem to be a sufficiently safe way of ensuring that sensitive data is protected. Specifically, NIH should consider requiring that datasets that are submitted and that contain sensitive information should be submitted to NIH at or above Level 2. This will help ensure that those accessing data that concerns sensitive or stigmatizing traits will only have access to non-individual level data.

Second, and building upon my first comment in this section, I believe that if datasets containing sensitive genomic information are restricted to those at specific levels, for example at or above Level 2 data, risk could be mitigated while still using the rapid access model. Still, I do not know to what extent these types of restrictions on data will impact an institution's ability to conduct primary and or secondary research. As the proposal indicates, if genomic summary results would risk divulging sensitive information, an institution may elect or that the dataset be provided only through controlled access. On the other hand, allowing data sharing via the rapid access model, but restricted to a Level 2 or greater would still allow for more secondary research to proceed while not risking sensitive information.

Lastly, since October 1, 2017 Certificates of Confidentiality have not been required for NIH research collecting sensitive data. As recent as this change is, it does not affect research that has already occurred. To what extent would NIH require research that occurred before the enactment of the 21st Century Cures Act, and the implementing policy that NIH promulgated to be released via rapid access? Would an institution still be able to determine what information would be made available? COCs are designed to protect an institution from having to disclose sensitive information, but if this is changed and they are no longer required would that change what information could potentially become subject to disclosure via rapid access?

4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies

There are some issues that I believe should be more thoroughly vetted and/or considered by NIH including: (1) The applicability of this policy to future research; (2) The applicability of this policy to the Freedom of Information Act (FOIA) more importantly, information gathered and shared that is publicly available, but potentially identifiable; and (3) The potential impact that the 21st Century Cures Act may have on implementation of rapid release as it relates to protecting personal information.

First, the GDS Policy specifically mentions that the policy applies to those projects being submitted to the January 25, 2015 receipt date, contract proposals on or after January 25, 2015, and for intramural projects generating genomic data on or after August 31, 2015. The same concern over the ways in which NIH intends to address informed consent for

those studies involving human participants exists for the way in which NIH proposes to address informed consent if it adopts rapid access. The GDS Policy strongly encourages investigators to obtain informed consent for “broad sharing and future research” purposes for research prior to January 25, 2015. Would “broad sharing and future research” include allowing for rapid release? It would seem that this would not need to be re-examined, and that if an investigator does go back to obtain modified consent forms for broad sharing and future research, would NIH interpret this to include rapid release so that ongoing research before January 25, 2015 could move forward?

Second, information submitted to an NIH repository become U.S. government records subject to FOIA. As a result they may be requested and NIH must comply with any requests, unless certain exceptions apply. As a general matter, and as is discussed in the proposal, data is more easily re-identified. While it is pretty clear that the NIH, and other agencies that may gain access to NIH repository information may be able to claim the exception under FOIA related to medical records (5 USC 552(b)(6)), does NIH have general concerns about the way in which sensitive human data should be managed in an age where re-identifying information is easier?

Lastly, on December 13, 2016 then President Obama signed into law the 21st Century Cures Act. This law sought to streamline data sharing and the informed consent process, especially in the context of genomic and human information. For example one significant change that came as a result of this law was Section 13444 that permits the remote access to protected health information by researchers. How does NIH intend to conform what is in this proposal to what Congress set out in the 21st Century Cures Act? Congress specifically acted to streamline and speed up genomic and personalized medicine research. I view rapid release as a serious way that NIH is committed to doing just that, but does NIH see any instance where this proposal would be inconsistent with that significant piece of legislation?

Additional Comment (attachment):

1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity.

The procedures that govern data sharing, or access procedures for the genomic summary results under the Genomic Data Sharing (GDS) Policy must be viewed through the lens of both the goals of the National Institutes of Health (NIH), as well as the more specific, but similar goals of any genome wide association studies (GWAS). Indeed, First it is important to consider the goals of the National Institutes of Health (NIH). Some of these goals include “foster[ing] creative discoveries, innovative research strategies [...] expand the knowledge base in medical and associated sciences in order to enhance the Nation’s economic well-being and ensure a continued high return on the public investment in research.” Second, and similar to the goals set by NIH are the goals must also be viewed in a more narrow, but relevant context of the goals of a genome-wide association study (GWAS) repository: (1) improve health; and (2) maximize public investment. The sharing of genome-based research with a broad audience will permit our nation’s scientific understanding of the human condition to grow and improve, and lead to the development of better technologies, while reducing the costs associated with the development of these technologies. However, these two ideas must be managed in the context of the applicable federal laws and regulations that govern and protect research participants’ privacy interests. For purposes of these comments, the GDS Policy applies to all large-scale data research projects funded by NIH like GWAS.

Presently, there are two tiers of access available: (1) unrestricted or open access; and (2) controlled access. From my perspective, this new proposal intends on providing a hybrid of the two – rapid access. Users will be required to agree to the terms set forth in a ‘click-through agreement.’ Concerning the degree to which access is granted, the rapid access and ‘click-through agreement’ must comport with the existing NIH framework that governs data sharing. While I share thoughts on the process concerning data submission in Question 2, the framework for how data is submitted needs to also be considered in conjunction with the degree to which access is granted, and what may need to be added to the ‘click-through agreement.’ Addressing the risks and benefits of rapid access is beneficial, but so is any action that is taken even at the very beginning when an NIH grant is awarded, e.g. What is included in the Genomic Data Sharing Plan that is submitted to NIH? What role will an institution’s IRB play in providing controls for ensuring that the genomic data that is generated is done so in accordance with the Common Rule, HIPAA, and other applicable NIH rules and industry standards.

First, NIH needs to consider reevaluating and updating their Data Use Certification (DUC) Agreements. It goes without saying that the purpose behind the rapid release of datasets, along with a shorter 'click-through agreement' is to accelerate the speed at which datasets may be requested. However, the DUCs that allow a principal investigator (PI) to request and subsequently use a genomic study dataset are designed in such a unique way so as to provide accountability and transparency so that the sensitive information that is contained in their research is protected. DUCs require that those who request access to controlled access datasets are going to use the information that they requesting responsibly. To reduce these to a 'click-through agreement' is somewhat concerning, though I do believe that it is possible to have an enhanced 'click through agreement' that would still incorporate the information already included in a DUC, but in a reduced format that could certainly be more user friendly. That being said, ensuring that the information included in DUCs is also incorporated into the 'click through agreement' is critical.

The 'click through agreement' as proposed requires that those seeking to use GWAS datasets are confirming that the user: (1) has reviewed the information resources describing the appropriate uses of genomic data, including genomic summary results; (2) will not attempt to re-identify or contact any individual or group within a study population; and (3) will promote scientific research or health through the use of the genomic summary results. Even still, these three components not begin to encompass the information that is contained in a DUC. Some recommendations that NIH may want include with the 'click through agreement' could be: (1) Providing specific details on security measures that the PI requesting the dataset will use; (2) What a PI requesting information would need to do with the information, or how access would be treated following the use of the information; (3) What oversight would be applied to PIs should they violate the DUC; and (4) How would an individual IC address the sharing of data that might contain particularly sensitive or potentially stigmatizing information on participating individuals. In summation, to what extent would the 'click through agreement' as proposed incorporate these additional requirements that already exist in DUCs? I believe that the information that a requesting institution makes in their data access request, and subsequent DUC is important to mitigate against improper data use.

Second, and in conjunction with those data management safeguards incorporated in at DUC, are the seven elements included as part of the NIH Code of Conduct for Genomic Data Use. The 'click through agreement' as presently proposed contains a couple of these elements. Namely, it requires that a PI will not attempt to identify or contact any individual in the study, but a review of, and incorporation of these seven principles reaffirm what is required as part of a DUC, and will ensure that those investigators who are requesting access to a dataset meet

some requirements more than what appear to be outlined in the proposal. It will avoid the risks of PIs misusing data, but still allow for some degree of faster access to the dataset.

2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity.

Generally the maintenance and subsequent protection of highly sensitive personal information has been a concern since 1974 when the Congress passed the 1974 Privacy Act. They recognized then, as subsequent laws and regulations reflect now, the high level of importance of securely storing personal information in this increasingly digital age. NIH should weigh the public good that comes as a result of genomic information versus the risk of having personal information maintained in the rapid access format. This arrangement must therefore necessarily incorporate the following considerations in the context of the Common Rule, and HIPAA: (1) IRB approval for the use of data as described in DUC and the extent to which protected health information (PHI) will be maintained so as to avoid HIPAA violations especially when a researcher is also an individual's health care provider; (2) The informed consent of individuals so that minimal risk is ensured with the new rapid access model, and 'click through agreement';

First, the Common Rule insists upon that the IRB approval process is an integral first step for any research project involving human subjects. IRBs are responsible for determining whether or not a research project can even move forward, and are required to include written procedures for the risks associated with a project that is supported by the federal project. This would relate to the collection of, but also the storage and maintenance of this information. As proposed the new rule would allow institutions that have particularly stigmatizing or sensitive personal information to submit Institutional Certificates that would include the requirements outlined by the Common Rule. Specifically, these certificates require that an institution have its IRB affirmatively approve the sharing and maintenance of the genomic data of its research, and that this approval include the way in which the data will be stored and disseminated. This presently includes the option for unrestricted or controlled access data sharing. I believe that the proposed alternative to allow institutions submitting particularly sensitive or stigmatizing information to further restrict their data from the rapid access model is appropriate. IRBs have been established as gate keepers for information, and allowing them the ability to "opt out" of allowing stigmatizing data to be shared is not only consistent with the Common Rule, but facilitate things administratively for NIH.

Second, while the NHGRI Workshop “Sharing Aggregate Genomic Data” concluded that there is minimal risk associated with loosened data access standards such as the proposed rapid access model, what is NIH capable of doing to ensure that it pose a minimal risk to participant data? Assumedly, the data that will be exchanged is de-identified which would already place the genomic information at a minimal risk level however, the proposed changes include systematically computed statistics to determine the risks associated with re-identification of information. These statistics will inform the NIH as to the risks that are associated with the storage of, and re-identification of sensitive personal data. As outlined in the proposal, these summary statistics will be available through the repository’s website and consequently public. My concern is how will NIH determine for purposes of deciding what degree of risk is proportional, and/or appropriate for the data being stored? What thresholds will be established, or will this be on a case-by-case basis? Referring to my first comment in this section, would NIH still be able to allow rapid access to information that is potentially sensitive, but of a very low risk of being re-identified? Further, a majority of the Common Rule restrictions are triggered if the level of risk for data is above a “minimal risk” level. For consistency purposes, does NIH propose to maintain the same definition for minimal risk as is included in the Common Rule definition and would NIH need to establish well-defined thresholds that clarify the risk proportionality discussed in the proposed changes?

Additionally, the Common Rule requires that a PI obtain informed consent for each human subject that meet specific requirements. While the newly finalized modifications to the Common Rule approved in 2017 make some minor changes to the way in which informed consent may be obtained, the Common Rule does require that participants be provided (1) A description of any foreseeable risks; and (2) How the confidentiality of a participants records will be maintained, amongst other requirements. To the extent that the proposal seeks to change the way in which access is granted for summary results, and that access is weighed in function to the risk associated with that data, how does the NIH anticipate modifying informed consent for subjects into the future? Also would NIH require that those institutions seeking to grant rapid access, provide information to participants about the systematically computed statistics associated with their data, and the way in which risk proportionality would be determined? One of the reasons for the revision of the Common Rule was to clarify the informed consent process, but I fear that this would not contribute to that goal. I feel that explaining the rapid access model could be confusing. Does NIH propose a way to accommodate this potential change to the way in which informed consent is obtained?

3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as

sensitive.

It is entirely appropriate under the new proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive. The objective of this rapid release model is to enhance the sharing of information for the advancement of public health while respecting, and protecting that information that is most susceptible to stigmatizing individuals. Providing the option for institutions to continue to specify that data can be provided only through controlled access is an important option that PIs and institutions still have under this update, and I believe that is important. NIH should also consider the following: (1) How will data sharing for sensitive data with the new rapid release model be affected if at all by the levels of data sharing; (2) Would this policy apply to both intramural and extramural research; (3) How would Certificates of Confidentiality (“COCs”) be addressed to potentially still maintain confidentiality, while allowing rapid access to data.

First, NIH provides general guidance on submitting data to NIH data repositories with regard to how data that undergoes different levels of data processing. Assumedly for the purpose of the primary and secondary research, the level to which data (e.g. Level 0, Level 1, Level 2, etc.) is processed is very important for privacy protection for those involved in human subjects, as well as the time at which the data is released. In the proposal it mentions that it is possible for an institution to specify how the data, specifically sensitive, and potentially stigmatizing PHI should be handled, and to whom should access be granted – either controlled or unrestricted access. Practically, how would this be administered through the Institutional Certification as the proposal indicates? Summary level data or secondary research would seem to be a sufficiently safe way of ensuring that sensitive data is protected. Specifically, NIH should consider requiring that datasets that are submitted and that contain sensitive information should be submitted to NIH at or above Level 2. This will help ensure that those accessing data that concerns sensitive or stigmatizing traits will only have access to non-individual level data.

Second, and building upon my first comment in this section, I believe that if datasets containing sensitive genomic information are restricted to those at specific levels, for example at or above Level 2 data, risk could be mitigated while still using the rapid access model. Still, I do not know to what extent these types of restrictions on data will impact an institution’s ability to conduct primary and or secondary research. As the proposal indicates, if genomic summary results would risk divulging sensitive information, an institution may elect or that the dataset be provided only through controlled access. On the other hand, allowing data sharing via the rapid access model, but restricted to a Level 2 or greater would still allow for more secondary research to proceed while not risking sensitive information.

Lastly, since October 1, 2017 Certificates of Confidentiality have not been required for NIH research collecting sensitive data. As recent as this change is, it does not affect research that has already occurred. To what extent would NIH require research that occurred before the enactment of the 21st Century Cures Act, and the implementing policy that NIH promulgated to be released via rapid access? Would an institution still be able to determine what information would be made available? COCs are designed to protect an institution from having to disclose sensitive information, but if this is changed and they are no longer required would that change what information could potentially become subject to disclosure via rapid access?

4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies.

There are some issues that I believe should be more thoroughly vetted and/or considered by NIH including: (1) The applicability of this policy to future research; (2) The applicability of this policy to the Freedom of Information Act (FOIA) more importantly, information gathered and shared that is publicly available, but potentially identifiable; and (3) The potential impact that the 21st Century Cures Act may have on implementation of rapid release as it relates to protecting personal information.

First, the GDS Policy specifically mentions that the policy applies to those projects being submitted to the January 25, 2015 receipt date, contract proposals on or after January 25, 2015, and for intramural projects generating genomic data on or after August 31, 2015. The same concern over the ways in which NIH intends to address informed consent for those studies involving human participants exists for the way in which NIH proposes to address informed consent if it adopts rapid access. The GDS Policy strongly encourages investigators to obtain informed consent for “broad sharing and future research” purposes for research prior to January 25, 2015. Would “broad sharing and future research” include allowing for rapid release? It would seem that this would not need to be re-examined, and that if an investigator does go back to obtain modified consent forms for broad sharing and future research, would NIH interpret this to include rapid release so that ongoing research before January 25, 2015 could move forward?

Second, information submitted to an NIH repository become U.S. government records subject to FOIA. As a result they may be requested and NIH must comply with any requests, unless certain exceptions apply. As a general matter, and as is discussed in the proposal, data is more easily re-identified. While it is pretty clear that the NIH, and other agencies that may gain access to NIH repository

information may be able to claim the exception under FOIA related to medical records (5 USC 552(b)(6)), does NIH have general concerns about the way in which sensitive human data should be managed in an age where re-identifying information is easier?

Lastly, on December 13, 2016 then President Obama signed into law the 21st Century Cures Act. This law sought to streamline data sharing and the informed consent process, especially in the context of genomic and human information. For example one significant change that came as a result of this law was Section 13444 that permits the remote access to protected health information by researchers. How does NIH intend to conform what is in this proposal to what Congress set out in the 21st Century Cures Act? Congress specifically acted to streamline and speed up genomic and personalized medicine research. I view rapid release as a serious way that NIH is committed to doing just that, but does NIH see any instance where this proposal would be inconsistent with that significant piece of legislation?

Submission ID Date: 10/19/2017

Name: Miriam Rodrigues

Name of Organization: Muscular Dystrophy Association of New Zealand

Type of Organization: Patient Advocacy Organization

Role: Patient Advocate

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

The benefits apparent from the rapid and open sharing of summary statistics from aggregate genomic data are that it brings tremendous benefit to the rare disease patient community via the scientific community, and the potential risks or harms of such sharing are largely theoretical. In our own experience of supporting researchers with the ExAC and gnomAD public variant frequency databases the benefits to academic, clinical and pharmaceutical scientists and therefore patients from sharing of aggregate data are substantial: The Broad report that the browsers have had over 10 million page views by over 200,000 unique users from 166 countries in the past three years, and have been used by diagnostic laboratories in the analysis of >50,000 rare disease families. This is significant.

Even greater value will arise as a result of broader sharing of aggregate statistics as empowered by the new policy.

One aspect of the new Genomic Summary Results Data Sharing Policy - the creation of a new tier of access, rapid-access, which requires a click-through agreement to gain access to summary statistics is of great concern to us and represents a huge risk. The concerns are summarized as follows: (1) Click-throughs make programmatic access to data-sets challenging; (2) they greatly complicate or prevent multiple important types of re-use of the data; and (3) they are highly unlikely to deter anyone with genuine malicious intent. Overall, we maintain that click-through agreements give the impression of extra protection, but actually do no good - and can even do non-trivial harm. ExAC and gnomAD do not and never have had click-through agreements, and we understand that no harm has ever come to participants as a result.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access.**

Risks and benefits may relate to participant protection issues and/or scientific opportunity

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

MDANZ urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible re-identification. The open access to variant frequency data through ExAC and gnomAD has been very important to research that has provided diagnoses to many patients with rare disease, including patients with rare neuromuscular disease in New Zealand.

Submission Date: 10/19/2017

Name: Janna Hutz

Name of Organization: Eisai, Inc.

Type of Organization: Biotech/Pharmaceutical Company

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Broad access to genomewide summary statistics will be of major benefit to drug discovery and development efforts. Drug targets that have been linked to the relevant therapeutic indication through human genetics have been shown to be roughly twice as likely to result in a drug approval (Nelson, et al. Nat Genet. 2015 Aug; 47(8):856-60.) As this relationship is increasingly appreciated, biotech and pharmaceutical companies have invested substantially in accessing and leveraging human genetic data to identify new drug targets and biomarkers and to repurpose molecules for new indications.

Genomic summary statistics from large, consortia-led studies of common disease are very useful in identifying loci that may be of therapeutic relevance, so it is essential that researchers at biotech and pharmaceutical studies are able to access this data. Even if a group seeks to develop a therapeutic for only one disease, it is still critical to access the worlds' summary statistics for all traits to understand potential on-target safety risks or alternate therapeutic indications that may be suggested by additional associations with the gene/variant of interest.

Being able to access this data through APIs and bulk download is therefore ideal. Click-through agreements for each dataset would make it difficult to allow for phenomewide assessments as described above. In addition to genomewide association study summary statistics, variant frequency data from ExAC/gnomAD is necessary for interpreting associations of interest, and programmatic access to those resources is important for my research as well.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I support the designation of some datasets as sensitive. However, it may also be worth considering the motivations of subjects who may be participating in such studies. It is

my understanding that subjects in understudied populations may be motivated to participate in research in the hope that it will lead to new therapies. It may therefore be important to ensure that genomewide summary data from these studies can still be requested, perhaps with additional security measures. Addressing the issue of sensitivity by making available only subsets of this data may limit use for drug discovery.

3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive

I support the designation of some datasets as sensitive and hope that the NIH will offer guidance on how to establish this. Declaring sensitivity based on a high risk of reidentification for certain populations seems appropriate.

4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies

A key limitation in industrial researcher's current ability to access such summary data is variation in investigators' understanding and interpretation of consent limitations associated with sharing data. I have, on a handful of occasions, been denied summary statistics that are otherwise available to academic researchers because one or more subjects in the dataset that was used to generate the summary statistics did not approve of commercial use of their sample or data. Many datasets available on the NIAGADS portal established by the NIA for dissemination of summary statistics for Alzheimer's disease and related phenotypes are subject to such a restriction; only non-profit entities can access these results. While I believe these restrictions may have been established in the best interest of the research subjects, I hope that the NIH can provide more clarity with respect to interpretation.

While I work for a commercial entity, my work and the work of my colleagues does not involve buying or selling individual-level data or summary statistics. Rather, we wish to use the knowledge yielded from these results to develop new therapies to improve patient's lives, and these may eventually be commercialized.

It is my personal opinion that summary statistics are results (no longer subject to restrictions from individual-level data), and that sharing these summary results with individuals in the pharma/biotech industry should not be limited by individual-level consent restrictions. Indeed, subsets of these results are frequently published in manuscripts that describe the top associations or that show a Manhattan plot, so it is unclear to me why genomewide results would be treated differently.

As part of the NIH's proposed new policy, explicit guidance from the NIH on access of summary statistics by researchers at for-profit entities would be immensely appreciated and impactful for the development of novel therapies.

These views are my own and are not representative of my employer.

Submission Date: 10/20/2017

Name: Samantha Barratt Ross

Name of Organization: Centenary Institute

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):

I am writing as an avid user of the ExAC and gnomAD databases. I use these resources to identify clinically relevant DNA variants in patients with inherited heart diseases.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction – there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Samantha Barratt Ross

Submission Date: 10/20/2017

Name: Joowook

Name of Organization: Guy's Hospital

Type of Organization: Healthcare Delivery Organization

Role: Medical Provider

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

I am writing as an avid user of the ExAC and gnomAD databases. We use these resources to provide clinical interpretation of genetic diagnostic test results. These resources are crucial for this.

I believe that the new "Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy" is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to our clinical practice.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open

access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

Sincerely,

Joo Wook Ahn

Submission Date: 10/20/2017

Name: Michele Mattioni

Name of Organization: Seven Bridges

Type of Organization: Biotech/Pharmaceutical Company

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through TCGA has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically.

Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including TGCA, TARGET, ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/20/2017

Name: Daniel MacArthur and Jessica Alfoldi

Name of Organization: Broad Institute of MIT and Harvard

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

We would like to applaud the NIH for moving in the right direction with its new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy”. The rapid and open sharing of summary statistics from aggregate genomic data brings tremendous benefit to the scientific community, and the potential harms of such sharing are largely theoretical. Our own experience with the ExAC and gnomAD public variant frequency databases has shown that the benefits to academic, clinical and pharmaceutical scientists from sharing of aggregate data are substantial: The browsers have had over 10 million page views by over 200,000 unique users from 166 countries in the past three years, and have been used by diagnostic laboratories in the analysis of >50,000 rare disease families. Even greater value will arise as a result of broader sharing of aggregate statistics as empowered by the new policy.

However, we are still very concerned by one aspect of the new Genomic Summary Results Data Sharing Policy - the creation of a new tier of access, rapid-access, which requires a click-through agreement to gain access to summary statistics. These concerns can be summarized as follows: (1) Click-throughs make programmatic access to data-sets challenging; (2) they greatly complicate or prevent multiple important types of re-use of the data; and (3) they are highly unlikely to deter anyone with genuine malicious intent. Overall, our position is that click-through agreements are a security fig leaf that gives the impression of extra protection, but actually do no good - and can do non-trivial harm. And we would like to emphasize that ExAC and gnomAD, along with other aggregate data sharing sites such as the Exome Variant Server, do not and never have had click-through agreements, and to the best of our knowledge no harm has ever come to participants as a result.

To explain those points in a bit more detail:

1. It is critical for summary statistic resources such as gnomAD that we allow access through programmatic interfaces (APIs) so that people can query them using software (e.g. pull out just a targeted slice of the data) and perform automated queries (e.g. pull out

the frequency of a specific variant when a user loads a different web page about that variant). Most implementations of click-through agreements will prevent or greatly complicate any form of programmatic access. There are possible technical workarounds, but all of them result in some kind of barrier to programmatic access.

2. Probably the single biggest obstacle created by click-through agreements is that they prevent or substantially complicate data re-use. Right now anyone can download the complete list of frequencies from gnomAD and load it up in another website, or use it to build other useful web services (the complete ExAC data set has been downloaded thousands of times). With any kind of click-through agreement they either couldn't do that at all, or would have to incorporate the same agreement in their usage policy, which may be incompatible with their proposed usage.

3. Most importantly, click-through agreements do nothing to prevent the types of usage that are most likely to be harmful. It is worth noting that ExAC and gnomAD have existed on the web for almost 3 years and been accessed more than 10 million times without us being aware of a single incident that has any risk of harming participants. The vast majority of users are simply interested in using the data in their research. The theoretical bad actor who is interested in malicious usage is extremely unlikely to be dissuaded by a click-through agreement, nor does a click-through agreement offer any real after-the-fact protection if a malicious actor decides to do harm.

We do agree that it would be responsible to allow data providers to keep summary statistics behind controlled access when they involve vulnerable populations or stigmatizing traits, so long as this is in accordance with the wishes of the population in question. Many apparently vulnerable populations, such as rare disease patients, are in fact strongly in favor of sharing data as broadly and openly as possible. We would urge decision-makers to consult carefully with populations before deciding on the protection their data requires.

In summary, while we applaud the decision to explicitly permit broad access to genomic summary statistics, we believe that click-through agreements will degrade or destroy programmatic access and data reuse, without having any meaningful effect on participant safety. Any policy that advocates for click-through agreements as a solution should spell out explicitly exactly what types of misuse the click-through will prevent, and should justify the non-trivial barriers to data usage that would result.

Signed,

Daniel MacArthur and Jessica Alfoldi

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/20/2017

Name: Ara Tahmassian

Name of Organization: Harvard University

Type of Organization: University

Role: Institutional Official

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

See attached

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):



Harvard University
Office of the Vice Provost for Research

Ara Tahmassian
University Chief Research Compliance Officer
ara_tahmassian@harvard.edu
vpr.harvard.edu

Richard A. and Susan F. Smith Campus Center
1350 Massachusetts Avenue, 836
Cambridge, MA 02138
t. 617.495.9797 f. 617.495.8051

October 20, 2017

VIA: <https://osp.od.nih.gov/gsr-rfi/>

Office of Science Policy
National Institutes of Health
6705 Rockledge Drive, Suite 750
Bethesda, MD 20892

Re: Proposed Update to the NIH Genomic Data Sharing Policy's Access Model for Genomic Summary Results.

On behalf of the President and Fellows of Harvard University ("Harvard University"), I write to thank you for the opportunity to comment on NOT-OD-17-110, Proposed Update to the NIH Genomic Data Sharing Policy's Access Model for Genomic Summary Results. We fully endorse the goal of this update to provide access to genomic summary results through methods proportional to the risks and benefits posed by this type of information. The new "rapid access" tier eliminates access barriers to aggregate data and this is a welcome change for the research community. Additionally, the new "rapid access" tier more closely aligns with the interests of research participants who wish to see the data collected using their samples broadly used to advance research into treatments for the diseases afflicting them.

The proposed changes the NIH is proposing regarding access to dbGaP data will help to streamline access for investigators seeking to use the data in large bioinformatic studies. It would make such information more readily available to folks who only require this aggregate level of data for their project and also would enable faculty who seek individual-level data to more quickly review and assess whether the summary results would actually meet their research needs prior to unnecessarily initiating the full controlled access process.

We do request one clarification to the final notice. Consistent with the expectations under the NIH GDS Policy, NIH expects that consent forms and the informed consent process for human genomic studies to clearly articulate the access plans for data and information generated through the study, including genomic summary results. In particular, the notice states that:

NIH expects consent processes and other information available to potential research participants to be transparent that participation in an NIH-supported study infers an acknowledgement that investigators may aggregate and analyze the data generated through the study. NIH expects that consent processes and other information explain that such analyses or other summaries of study information (including genomic summary results) will be shared in the scientific literature or through other public scientific resources, such as data sharing resources that provide broad or unrestricted access to the information.

We assume this consent requirement is limited to prospective collections and not applicable to genomic summary results that were previously uploaded to dbGaP and now eligible for the rapid access tier. However, this is not clear in the notice. We request that this be made clear in the final notice, to avoid any confusion as to the applicability of this requirement.

Thanks again for the opportunity to comment.

Sincerely,



Ara Tahmassian, P.h.D.

Submission Date: 10/20/2017

Name: Larry Jacques

Name of Organization: Sault Ste. Marie Tribe of Chippewa Indians

Type of Organization: Government Agency

Role: Government Official

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Institutions who are submitting datasets on Native American people must include in their data sharing plan that any release of information to NIH must first pass through a Tribal Institutional Review board that specifically includes a section for the protection of Tribal Communities. If this happens before entering into the NIH data repositories, this could result in the mitigation of some risk.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

The maintaining of genomic summary results on a Tribal population with unlimited and indefinite use is dangerous to the ongoing trust relationship between Tribal populations and the National Institute of Health. An ongoing relationship requires that we build trust methodically over time, and by building on that trust NIH can have increased participation and eventually provide more specific information for our smaller populations. To build trust we have to build into the system the ability to allow for future determination of if the genomic information collected from tribal populations is causing harm. If that is the case, any member of our population should ALWAYS be able to pull that information out of a data collection even if that results in the statistical invalidation of a study. Having a system that supports the ability to pull information back is necessary to build trust and to build a feeling of control which is often difficult to have when interacting with large institutions such as NIH. We want to see this institution that at its core is actively establishing trust so that we can someday realize the full benefit of NIH research for our Tribal populations.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

The lack of Native American Investigators means that it is unlikely that we will get a Native American Investigator working with our datasets. To rely on the policy verbiage of having the (most likely non-tribal) investigator to be "expected to indicate" whether or not their research falls under the scope of GDS policy seems a little concerning. Again the utilization of a Tribal Institutional Review Board may assist in getting more than one set of eyes on the protection of sensitive populations.

4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies

In general, the response time that you are allowing for is inadequate for most tribes to identify the issue, to staff the issue, and to establish a response. To be more honest about the protection of sensitive populations it would be great to give more time for those who are responsible for defending small populations to become aware and to respond to these proposed policy changes.

Additional Comment (attachment):

NIH Request for Comment on Proposed Update to the NIH Genomic Data Sharing Policy's Access Model for Genomic Summary Results

Comments from Sault Ste. Marie Tribe of Chippewa Indians

1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity

Response:

Institutions who are submitting datasets on Native American people must include in their data sharing plan that any release of information to NIH must first pass through a Tribal Institutional Review board that specifically includes a section for the protection of Tribal Communities. If this happens before entering into the NIH data repositories, this could result in the mitigation of some risk.

2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity

Response:

The maintaining of genomic summary results on a Tribal population with unlimited and indefinite use is dangerous to the ongoing trust relationship between Tribal populations and the National Institute of Health. An ongoing relationship requires that we build trust methodically over time, and by building on that trust NIH can have increased participation and eventually provide more specific information for our smaller populations. To build trust we have to build into the system the ability to allow for future determination of if the genomic information collected from tribal populations is causing harm. If that is the case, any member of our population should ALWAYS be able to pull that information out of a data collection even if that results in the statistical invalidation of a study. Having a system that supports the ability to pull information back is necessary to build trust and to build a feeling of control which is often difficult to have when interacting with large institutions such as NIH. We want to see this institution that at its core is actively establishing trust so that we can someday realize the full benefit of NIH research for our Tribal populations.

3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive (limit: 8000 characters)

Response:

The lack of Native American Investigators means that it is unlikely that we will get a Native American Investigator working with our datasets. To rely on the policy verbiage of having the (most likely non-tribal) investigator to be "**expected to indicate**" whether or not their research falls under the scope of GDS policy seems a little concerning. Again the utilization of a Tribal Institutional Review Board may assist in getting more than one set of eyes on the protection of sensitive populations.

4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies (limit: 8000 characters)

Response:

In general, the response time that you are allowing for is inadequate for most tribes to identify the issue, to staff the issue, and to establish a response. To be more honest about the protection of sensitive populations it would be great to give more time for those who are responsible for defending small populations to become aware and to respond to these proposed policy changes.

Thank you.

Larry Jacques

Director of Strategic Planning

Planning and Development Department

Sault Ste. Marie Tribe of Chippewa Indians

523 Ashmun St.

Sault Ste. Marie, MI 49783

Office PH. 906.635.6050 ext 26049

Submission Date: 10/20/2017

Name: NHGRI-EBI GWAS Catalog

Name of Organization: EMBL-EBI

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

We support the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” but not the section that proposes to impose any click-through agreements. In our experience at EMBL-EBI, click-through agreements restrict data sharing and reuse and there is little evidence to show they prevent harm. The GWAS Catalog would like to become the principal host of summary statistics and to have these data freely available for reuse and download. Click-through would prevent programmatic access to our summary statistics and make it impossible for bulk querying or automated querying of the data sets. We have many other resources that incorporate our data but this secondary distribution would not be possible with a click-through agreement.

We urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible re-identification.

Submission Date: 10/20/2017

Name: Pamela Herd

Name of Organization: University of Wisconsin-Madison

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

I strongly believe these summary statistics, produced with federal funding, should be far more accessible to the broader research community. There is little reason to be concerned about issues of confidentiality, and the profound broader benefits in terms of the additional scientific work that can be done with better access to these data far outweighs any other issues.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/20/2017

Name: Daniel MacArthur and Jessica Alfoldi

Name of Organization: Broad Institute of MIT and Harvard

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

As an extra note, we would like to comment on the 'Informed Consent' section of the Proposed Update to Genomic Summary Results Access. If the intent of the policy is to encourage future consent forms to explicitly inform participants that their deidentified data may be released publicly in aggregate form, then we wholeheartedly support it. However, if the policy is intended to be applied retroactively, restricting the release of summary statistics to those studies with such language already in their consent forms, we would note that this would result in the prohibition of all currently available summary statistic based resources. We suggest that a clarification on this point would be helpful.

Submission Date: 10/20/2017

Name: Lisa Nichols

Name of Organization: Council on Governmental Relations

Type of Organization: Nonprofit Research Organization

Role: Member of the Public

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):



October 20, 2017

Via: <https://osp.od.nih.gov/gsr-rfi/>

Office of Science Policy
National Institutes of Health
6705 Rockledge Drive, Suite 750,
Bethesda, MD 20892

Re: Proposed Update to the NIH Genomic Data Sharing Policy's Access Model for Genomic Summary Results

The Association of American Universities, the Association of Public and Land-grant Universities and the Council on Governmental Relations appreciate the opportunity to comment on NOT-OD-17-110, Proposed Update to the NIH Genomic Data Sharing Policy's Access Model for Genomic Summary Results. Our associations endorse the agency's efforts to improve access to, and promote broad sharing of, aggregate genomic data held in NIH repositories (e.g., the Database for Genotypes and Phenotypes), in a manner that is proportional to the risks and benefits posed. We further support the use of a "rapid access" tier that requires users to affirm agreement with a statement regarding responsible use of the information. With respect to "potentially stigmatizing traits" and "increased privacy risk or heightened risk of group harm" the nature of the data should largely preclude such risks. However, we support the approach of providing institutions submitting data sets the option to indicate, as part of their Genomic Data Sharing Plan and Institutional Certification, that summary results should be provided only through controlled access as appropriate.

We would suggest that the implementation plan with respect to datasets submitted to or accessible through designated repositories prior to the effective date include a common process for notifying funding Institutes or Centers if summary data should be maintained in controlled access, and that the process not require re-certification of the submitted or currently accessible data. We also request that NIH clarify that the consent requirements detailed in the notice are limited to prospective collections.

Thank you again for the opportunity to comment.

About the Signatory Associations

The Association of American Universities is an association of 60 U.S. and two Canadian preeminent research universities organized to develop and implement effective national and institutional policies supporting research and scholarship, graduate and undergraduate education, and public service in research universities. The Association of Public and Land-grant Universities (APLU) is a research, policy, and advocacy organization with a membership of 235 public research universities, land-grant institutions, state university systems, and affiliated organizations in the U.S., Canada, and Mexico, that is dedicated to strengthening and advancing the work of public universities. The Council on Governmental Relations (COGR) is an association of over 190 leading research universities and affiliated academic medical centers and independent research institutes. COGR concerns itself with the impact of federal regulations, policies, and practices on the performance of research conducted at its member institutions.

Submission Date: 10/20/2017

Name: James Lawlor

Name of Organization: HudsonAlpha Institute for Biotechnology

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I am concerned that the proposed click-through agreement mechanism will unnecessarily limit programmatic access to available data, which is critical in order to work with genomic-scale data in an efficient and reproducible fashion. I am skeptical that a click-through agreement would halt improper use of the data; it also seems unlikely that a click-through agreement would be effective in reinforcing the standards of ethical and allowable use of the data--I believe most readers can easily identify numerous similar click-through agreements they have personally bypassed without careful attention. I think the greater scientific community would be better served by implementing the same guidance in a non-obstructive manner. In short, I share the concerns outlined in the following article from Dr. Daniel MacArthur:

<https://macarthurlab.org/2017/10/10/response-to-proposal-to-update-data-management-of-genomic-summary-results-under-the-nih-genomic-data-sharing-policy/>

This comment represents my views as a practicing computational biologist; however, it should not be construed to represent an official institutional position nor representative of the views of any other employees or associates of my institution.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/20/2017

Name: Or Zuk

Name of Organization: Hebrew University of Jerusalem

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I develop statistical and computational methods for interpreting genetic data, with a focus on human genomics.

I am an avid user of the ExAC and gnomAD databases, to which I apply many of my methods to gain insights into human evolution and complex traits.

I believe that the new “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy” is a step in the right direction - there is no evidence that controlled-access of summary statistics prevents any harm to participants, and the open access to variant frequency data through ExAC and gnomAD has been very important to my research.

However, I am concerned about the proposal to create a new rapid-access category for summary statistics that would require the use of click-through agreements. These agreements make it difficult to reuse summary statistics and to access data programmatically. Most importantly, there is no evidence that they prevent harm to participants. A wide variety of summary statistics have been publicly available without click-through agreements for many years, including ExAC and gnomAD, and no harm of any kind has ever been done to any participant whose data is aggregated in those summary statistics in that time.

I urge the NIH to modify this proposal, and to designate summary statistics as open access, with the exception of communities and populations who believe that they are especially vulnerable to harm from possible reidentification.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/20/2017

Name: Allison Provost

Name of Organization: Cohen Veterans Bioscience

Type of Organization: Nonprofit Research Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Cohen Veterans Bioscience supports timely sharing of genomic summary results to accelerate discovery. Genomic analyses have propelled advances in both clinical care and basic science approaches over the last decades. Data availability and accessibility are critical bottlenecks for this acceleration. Identification of causal genetic variants in disease allows for the discovery of new, improved treatments and next generation diagnostics. Ultimately, researchers seek to improve patients' lives through these discoveries, but unrestricted sharing may put some patients at risk since it may be possible to identify whether a patient has participated in a study based on population statistics. Thus, data-sharing should occur with the patients' best interests in mind. Balancing the tension between wide-sharing and patient protection is key, particularly for stigmatized disorders like psychiatric disorders.

Submission Date: 10/20/2017

Name: Prof Mark Caulfield

Name of Organization: Genomics England

Type of Organization: Government Agency

Role: Institutional Official

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Genomics England broadly welcomes the direction of the proposed NIH Genomic Data Sharing Policy. Genomics England puts participants at the heart of the 100,000 Genomes Project. We also recognize the value in researchers being able to nimbly access and use data in order to further our understanding of the human genome and ultimately develop better treatments for patients in the future. We use gnoMAD in the analysis of all whole genomes in the 100,000 Genomes Project. This has allowed us to return reports on 5,000 genomes to the NHS, offering 20 to 25% actionability to people who have been waiting, in some cases, up to 28 years for a diagnosis. We cannot underscore sufficiently the value to our patients and ourselves of a versatile and accessible resource such as gnoMAD. All similar healthcare programmes will very likely need to use this highly valued resource. With projections of 60 million whole genome sequences in rare disease in the next decade many healthcare professionals will wish to access these data. It is with patients and these aims in mind, that we urge NIH to carefully examine the ‘click through’ agreement in detail, particularly how this will affect the use of a wide variety of summary statistics. We respect the need to protect sensitive data but would ask NIH to ensure it is proportionate and will not only protect participants but also not unduly hinder research or healthcare opportunities in the future that use aggregate data and ultimately aim to benefit humanity worldwide.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/20/2017

Name: National Society of Genetic Counselors

Name of Organization: National Society of Genetic Counselors

Type of Organization: Professional Org/Association

Role: Patient Advocate

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):

October 20, 2017
National Institutes of Health
Office of Science Policy
6705 Rockledge Dr., #750
Bethesda, MD 20817

Re: Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy

The National Society of Genetic Counselors (NSGC) appreciates the opportunity to comment on the National Institutes of Health's (NIH) Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy.

NSGC is the leading voice, authority, and advocate for the genetic counseling profession, which has grown to more than 4,200 certified genetic counselors. Genetic counselors are healthcare professionals with accredited specialized Master's degrees that focus on medical genetics and counseling. They provide information and support to individuals and families concerned with genetic disorders or birth defects, at risk for a variety of inherited conditions, or seeking genomic evaluation services. Their primary roles are to procure and interpret family and medical histories, identify at-risk individuals, explain genetic inheritance and natural history, quantify chance of occurrence and recurrence, review available testing options, facilitate decision-making regarding testing, provide results interpretation, and provide post-test counseling.

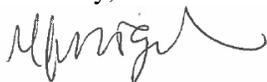
While open access genomic data speeds genetic research with the goal of improving the lives of individuals with genetic conditions and the general population, limitations to open access are necessary to protect the privacy of and respect the informed consent of research participants. NSGC recognizes the complexities of this proposal and requests that the NIH consider the following operational recommendations for any implementation.

1. If the NIH adopts the proposed access-model changes, research consents should explicitly indicate the intent to share data publicly through the NIH's open- or restricted- access databases. NSGC encourages the NIH to create verbiage for researchers to use in informed consent documents that discusses the risks and benefits of publicly shared data in both types of databases. When creating this verbiage, NSGC encourages the NIH to consult with genetic counselors who have expertise in the recruitment and consent of research participants.
2. It is important to recognize that pleiotropy, scientific advancements, or change in cultural norms could alter the perceived sensitivity of a particular trait. NSGC strongly recommends that if the NIH implements this new policy it a) directs research institutions to be clear and transparent about these issues in the informed consent document and b) ensures the individual obtaining consent is familiar with these issues and can explore them with the research participant.
3. The ethical standards for classification of this data will need to be above reproach. Classifying organizations or individuals will be entrusted to make unbiased decisions about which category the data falls into. Since there is an industry-wide incentive to classify data as "non-sensitive," it will be essential to be thoughtful and principled in this decision. An organization such as an Institutional Review Board (IRB), which vets for conflicts of interest in its members, should have a process for evaluating studies, and should meet regularly to classify and (if necessary) reclassify previously submitted studies. New developments will force classification changes for an indefinite period, therefore institutions and the NIH need transparent and expedient processes to transfer data from open to restricted access, and vice versa. The NIH's commitment to this process will need to be indefinite and ongoing.

4. NSGC is concerned about the data that is already in the dbGaP Database. Data in a controlled- access database should not be moved to an open-access database unless research participants explicitly consented to public sharing or the IRB that granted study approval approved the public sharing of data. If data were obtained without consent for open-access sharing or appropriate IRB approval, NSGC recommends that it remains in a restricted access database.

NSGC appreciates the opportunity to comment on the NIH's proposed genomic data sharing policy and looks forward to future opportunities to collaborate.

Sincerely,



Mary E. Freivogel, MS, CGC
President
National Society of Genetic Counselors
2025 M Street NW, Suite 800
Washington, DC 20036

Submission Date: 10/20/2017

Name: Steven Buyske

Name of Organization: Rutgers University

Type of Organization: Not Applicable

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

As others have written, it's very difficult to see harm coming from the posting of summary data under any but the most artificial of circumstances. In any case, having click-through agreements would not deter a bad actor; it would only allow the NIH to claim it they weren't responsible.

Submission Date: 10/20/2017

Name: Denise A Dillard

Name of Organization: Southcentral Foundation

Type of Organization: Other - Tribal health organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

If the workgroups who met prior to this proposal in 2012 and 2016 did not include members of the public including tribal members as well as tribal leaders, the risks of broad access have not been fully delineated. Although it sounds like consent will be considered and risks to indigenous populations are recognized, many NIH program officers and scientific reviewers do not seem aware of appropriate exceptions to widespread data sharing and continue to push submissions which are in direct contradiction to what most tribes require as sovereign nations. In addition to considering individual consent, tribal approval should be obtained for broad access to summary results involving tribal members in order to ensure adequate protection against group harms.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/20/2017

Name: Paul Flicek

Name of Organization: European Molecular Biology Laboratory, European Bioinformatics Institute

Type of Organization: Nonprofit Research Organization

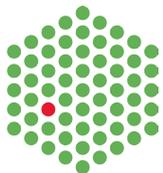
Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):

EMBL-EBI



European Molecular Biology Laboratory
European Bioinformatics Institute
Wellcome Genome Campus
Hinxton, Cambridge CB10 1SD
United Kingdom



Office of Science Policy (OSP)
National Institutes of Health
6705 Rockledge Drive, Suite 750
Bethesda, MD 20892
United States of America

20 October 2017

RE: Response to Proposed Update to the NIH Genomic Data Sharing Policy's Access Model for Genomic Summary Results

To all it may concern:

I am writing in my capacity as leader of the Ensembl project based at the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) based near Cambridge, England. Ensembl is one of the world's leading sources of genome information and a central aggregation point for genomic data. Ensembl software is fully open source and our data is freely available through our web browser at www.ensembl.org; our Perl and REST APIs; our BioMart data mining platform serving direct queries and BioConductor access via biomaRt; and direct downloads of all data. Ensembl's Variant Effect Predictor (VEP) is the premier open source variant annotation tool. VEP is actively supported and is used by thousands of companies, research groups and hospitals.

Ensembl's web interface supports over five million user sessions annually and many millions more—which are not captured by analytics—access Ensembl data via our APIs, tools and data incorporated programmatically or directly into hundreds of other bioinformatics resources. More than half of Ensembl's usage at our main sites is for human. As a reference resource for genomics, Ensembl aims to provide the most general and most useful data resources in a consistent form. We already incorporate genomic summary results from ExAC, gnomAD, TopMed, UK10K and other projects. These data are highly used in many contexts and are deeply embedded into the Ensembl ecosystem and tool set described above.

Since Ensembl's first release in 2000, our policy has always been to have as few barriers as possible: no limits are put on the data provided by the project, no Ensembl user pays a license fee, no data in Ensembl requires click-through access. We strongly believe Ensembl's openness has significantly contributed to the genomic revolution.

We understand and support the idea that human genomic data must be treated appropriately and so care must be taken when either erecting or removing barriers to access especially for vulnerable study populations where additional protection may be more easily justified. It is also possible that datasets arising

from vulnerable study population will have lower user demand than more general reference data sets and so may not be targets for incorporation to Ensembl and other resources. Regardless, accessing human genomic data remains more challenging than it should be in many cases and there is room for significant improvements to the process of controlled access for both summary and individual level data.

We are extremely supportive of the concept of further increasing access to genomic summary data. We believe that there are many benefits to doing so and have neither observed nor received any reports of harm associated with this type of access. However, we do not support the use of click-through licenses for genomic summary data and include in this response some of the potential consequences for the Ensembl platform should such a policy come into force.

1. Data removal vs. costly retrofitting

If the proposed click-through policy were immediately implemented, we would be forced to remove any summary data subject to the policy while we considered how, or if, to support the policy. Although adding a check box to the web site is relatively easy, Ensembl's power arises from deep and consistent integration across multiple tools including our APIs, VEP and BioMart. These tools include data access points that have been created over the project's history and rapidly retrofitting a seamless user interface into our system would be neither easy nor cheap.

Moreover, our archive websites and related resources, which stretch back five years for websites and longer for API and database access, are no longer subject to active development. Retrofitting this code would be extremely costly. The alternative, closing access to these data and websites, would disrupt on-going research projects and impair reproducibility of already published results. Unlike a commercial company we cannot hire several new developers for a year with the promise of greater profits after the development is complete. Our tools and services are free.

2. Variable international applicability

Scientific research is international. We are uncertain about the mechanism of enforcement for click-through agreements for researchers outside the United States. We also recognise that researchers using these data in various jurisdictions may not have the contractual authority to agree to the click-through license and this could limit their use of Ensembl and any other resource incorporating such data.

3. Challenges for data integration from various sources

The NIH does not make policy in a vacuum. Many funders and policy makers worldwide are likely to follow the lead of the NIH and we believe that it will be unlikely that more than a few will have less restrictive policies than what is enacted by the NIH. This is especially true for those organisations that may be

less inclined to share data for wide variety of reasons.

Ensembl currently supports summary genomics datasets funded by the NIH and the Wellcome Trust, but we expect that this list will grow considerably over the next several years as more sequencing takes places across the world. Supporting a growing number of click-through agreements would be unfeasible. In fact, we do not believe that statements from each funder or data source could (or should) be accommodated on every page providing summary data. **Instead, we urge the NIH to work with others around the world to create a Code of Conduct for use of summary genomic data.**

Sincerely,

A handwritten signature in black ink, appearing to read 'Paul Flicek', with a stylized flourish at the end.

Paul Flicek DSc
EMBL Senior Scientist
Wellcome Investigator
Team Leader, Vertebrate Genomics
Head of Genes, Genomes and Variation Resources

Submission Date: 10/20/2017

Name: Hae Kyung Im

Name of Organization: University of Chicago

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I agree with the click through agreement as long as it is not detrimental to usability. It should only require it once per device (computer, smart phone, etc). Also there should be a key management solution that ensures that programmatic access is seamlessly allowed. Download of full set of summary results should also be allowed.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Allowing exceptions of this sort has the potential to substantially decrease the summary results available publicly and will stifle scientific progress.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

I do not agree with this option.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

There should be enforcement of deposit of summary results as soon as a publication is accepted. This can be implemented using the same mechanism as used for PMC registration.

Also, summary results generated by the authors of the study will be more useful than the ones generated in a semi automated fashion by the NIH. The authors of the study are the best qualified people to execute the analysis.

Submission Date: 10/20/2017

Name: Jennifer Hall

Name of Organization: American Heart Association

Type of Organization: Professional Org/Association

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Potential risks associated with this change in policy of providing a rapid access mechanism to genomic summary level data with the associated click-through agreement are thought largely to be centered around participant protection issues.

The increased potential risk in data privacy and protection, although thought to be minimal, to patients/consumers and members of our communities needs to be communicated. The process in which we communicate these changes needs to be careful, attentive, and include a dialogue to be sure we listen to concerns.

At a strategic level, the change in policy of providing rapid access to genomic summary level data shifts the onus first to the clinical and research team to take increased responsibility and care in the education and consent process with participants. The change in policy also shifts the balance to each community to take an active role in educating their members/consumers/ patients and helping each member of the community understand what they agree to and the risks associated with signing such an agreement. At first blush, this shift in balance may seem unfair or misplaced. With time, training and guidance, the potential risks may help improve the understanding of consumers and the overall well-being of members of our community.

Finally, one hopes that change in policy and risk management also brings opportunities to accelerate new systems that advance patient protection, data privacy, data tracking, and security.

Potential benefits associated with this change in policy include increased scientific opportunity for all researchers, and increased time spent on data analysis vs. tracking down datasets. A potential benefit to consumers/patients/and members of the community is scientific acceleration towards cures and discoveries that inform disease prevention strategies.

2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity

A potential risk associated with this stipulation includes the grant review committee's willingness, expertise, and attention given to this section in applications as well as the potential uneven nature of different grant review committees. We suggest an overall process/committee that educates the grant reviewers in reviewing this section or stipulates an NIH staff review.

The benefit to the consumer/participant greatly outweighs the potential risks. Overall this is an important component of the plan that will likely need to be reworded as data continue to evolve.

3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive

The proposal is appropriate for institutions submitting study data under NIH GDS Policy.

4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies

Overall, the American Heart Association supports the change in policy for NIH-funded studies to adapt a rapid-access process to genomic summary level data. We support the stipulation around submitting institutions including sensitive information for studies in controlled access.

Submission Date: 10/20/2017

Name: Joyce Tung, David Hinds, Adam Auton, Sarah Elson

Name of Organization: 23andMe, Inc.

Type of Organization: Biotech/Pharmaceutical Company

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

At 23andMe we strive to maximize the utility of published datasets for use by the research community. We have developed policies and processes that streamline access to summary statistics from published GWAS, while also minimizing the risk of re-identification of individual research participants. We recognize that, especially with increasing sample sizes of genomic datasets, there are limits to feasibility of re-identification attacks (e.g., as described in Homer N et al., PLoS Genet. 4(8):e1000167, 2008). We support evaluating a risk-based approach that considers feasibility of attack and sensitivity of information that could be revealed, as opposed to a uniform level of controlled access for all aggregate genomic datasets.

23andMe has intentionally developed policies that may be more conservative than those of the NIH and academic institutions. We maintain a continuing relationship with our research participants, and it is a high priority to ensure customer privacy and security of customer information. A breach - even just a proof of concept attack - using 23andMe data could threaten our ability to continue to conduct research and/or our commercial viability. Due to 23andMe's public prominence, we may be more vulnerable to a proof-of-concept attack.

Part of the rationale underlying the current NIH proposal for more open access to genomic datasets was that the potential for harms was largely theoretical, and there have not been instances of successful re-identification attacks.

A successful attack requires access to an individual's genetic information, and inclusion of that individual's data in a public GWAS dataset. To date, there have been few sources of individuals' genetic information. Given the increasing numbers of individuals who have obtained raw genetic data through direct-to-consumer (DTC) testing (e.g., over two million people have now been genotyped through 23andMe), there may be increasing opportunity for a re-identification attack in the future. Similar opportunity for attack may exist when and if the All of Us initiative returns raw data to participants.

Whether via a click-through agreement and/or other mechanisms, we endorse policies that provide recourse and enforce substantive sanctions in the event of violations of the data access terms.

2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity

As described above, risk of re-identification attacks persists, and there is the possibility that there may be new mechanisms for attacks in the future that are not yet contemplated. Given the greater potential for harm to individuals represented in aggregate genomic datasets containing potentially stigmatizing or sensitive information, we believe that such datasets should be maintained as controlled access by default, with risk-based approaches to specific datasets.

3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive

We agree with the appropriateness of individual institutions indicating whether specific datasets should be designated as sensitive. Particularly in the absence of existing federal guidelines or consensus in the field about what constitutes sensitive or potentially stigmatizing information, individual institutions, whose personnel are most familiar with their specific study populations, are in the best position to understand these populations' particular concerns.

4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies

Submission Date: 10/20/2017

Name: Maria Kannu

Name of Organization: University Health Network

Type of Organization: Healthcare Delivery Organization

Role: Institutional Official

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Existing consents do not cover this kind of open sharing. It also means any new consents will require yet more wording to allow this, and, there does not appear to be a way to opt out and have this sharing turned off for an individual because they want to opt out (it breaks our chain of what we can say is mandatory vs optional).

Secondly, many of these disease aggregate databases that have been around often include phenotype and ancilliary data enough to zero in on someone. There is no technical description of what is allowed to be assembled together and shared. We need this in writing to make sure we don't back-door our data to our patients.

Submission ID Date: 10/20/2017

Name: Erin Ware

Name of Organization: University of Michigan

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Broad access to genomic summary results is essential to translate the large investment by NIH in genotyping and genome-wide association studies (GWAS) to benefit human health. Published summary results, often only genome-wide significant, or genome-wide suggestive variants, are important findings. But they are the tip of the iceberg. Important and useful information from GWAS/meta-analysis are contained in summary results for less-significant variants, e.g. for studies of gene-by-environment interaction, pleiotropy, pathway, and network analyses.

Having a limited set of broadly available results (genome-wide significant/suggestive only) limits the comparison of “best methods” for summarizing results across available GWAS. New methods - including those related to polygenic scores and multi-trait analysis of GWAS (MTAG) - require access to full summary statistics to produce robust estimates. A great deal of information is lost below these arbitrarily chosen thresholds for publication that are not consistent across studies. Broad, rapid release and access would increase scientific reproducibility and innovation within population genetic analyses.

As other “-omic” data become increasingly available, such as epigenome-wide association study data, they should be treated in a similar way.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

Sensitive data classification - Keeping sensitive data protected and abiding by strong research ethics are our highest priority. However, summary statistics from GWAS are not sensitive data. They cannot be used to identify participants. GWAS consortia producing

summary statistics are composed of hundreds of thousands of individuals from potentially dozens of studies. The identifiability of a single individual from summary results of these mega-meta-analyses is impossible. Attempts to identify whether an individual is included in a cohort (per Homer et al. 2008) require access to individual-level genomic data and a sufficiently representative set of allele frequencies. While this is hypothetically possible, to then use that information to identify that participant's individual level study data would be in violation of the data use agreement signed to gain access to the study dataset. Requiring investigators to affirm agreement with a statement regarding responsible use of the information in order to gain access to genomic summary statistics is a sufficient level of protection given the potential and hypothetical risk of identification

(<https://www.genome.gov/pages/policyethics/genomicdata/aggdatabereport.pdf>, see finding 6). Gaining access to individual-level genomic data should be regulated and controlled, access to genomic summary results should not.

Summary statistics are also not subject to restrictions of use related to participant consents. Participants in studies may consent to use of their data only for a specific research question. But summary statistics are not participant data. Moreover, selected summary statistics - those deemed statistically significant or nearly so - are already being published. It is not clear why results that fall below some arbitrary statistical cutoff should be treated differently. No protections are gained by withholding the full set of genomic results.

4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies

Speed of access to summary statistics - With the speed at which the genomic field moves, it is extremely important to have timely access to full summary statistics. We strongly urge that the availability of summary statistics to coincide with publication. This would not require analyses beyond those for publication, are technically part of the analyses for publication (usually not shown for space), and thus should be available concurrently.

Contents of summary statistic release - For maximum use and productivity, the released summary statistics should be standardized as much as possible. Our recommendation is that the summary statistics ought to contain the following: information on the genome build, strand alignment, SNP name, direction of effect, effect size, standard error, p-value, Allele 1 (effect allele), Allele 2 (non-effect allele), chromosome, position, effect allele frequency for all SNPs from the initial meta-analysis (e.g. discovery phase, phase 1), and additionally a release of any SNPs carried forward in the replication phase (e.g. replication, phase 2, joint analysis, etc.).

Additional Comment (attachment):

We are writing in response to the National Institutes of Health RFI seeking comments regarding the proposed update to the access procedures for genomic summary results under the Genomic Data Sharing (GDS) Policy ([NIH Guide Notice NOT-OD-17-110](#)).

Broad access to genomic summary results is essential to translate the large investment by NIH in genotyping and genome-wide association studies (GWAS) to benefit human health. Published summary results, often only genome-wide significant, or genome-wide suggestive variants, are important findings. But they are the tip of the iceberg. Important and useful information from GWAS/meta-analysis are contained in summary results for less-significant variants, e.g. for studies of gene-by-environment interaction, pleiotropy, pathway, and network analyses.

Having a limited set of broadly available results (genome-wide significant/suggestive only) limits the comparison of “best methods” for summarizing results across available GWAS. New methods – including those related to polygenic scores and multi-trait analysis of GWAS (MTAG) – require access to full summary statistics to produce robust estimates. A great deal of information is lost below these arbitrarily chosen thresholds for publication that are not consistent across studies. Broad, rapid release and access would increase scientific reproducibility and innovation within population genetic analyses.

As other “-omic” data become increasingly available, such as epigenome-wide association study data, they should be treated in a similar way.

In this response we would like to address several specific issues.

Speed of access to summary statistics - With the speed at which the genomic field moves, it is extremely important to have timely access to full summary statistics. We strongly urge that the availability of summary statistics to coincide with publication. This would not require analyses beyond those for publication, are technically part of the analyses for publication (usually not shown for space), and thus should be available concurrently.

Contents of summary statistic release - For maximum use and productivity, the released summary statistics should be standardized as much as possible. Our recommendation is that the summary statistics ought to contain the following: information on the genome build, strand alignment, SNP name, direction of effect, effect size, standard error, p-value, Allele 1 (effect allele), Allele 2 (non-effect allele), chromosome, position, effect allele frequency for **all** SNPs from the initial meta-analysis (e.g. discovery phase, phase 1), and additionally a release of any SNPs carried forward in the replication phase (e.g. replication, phase 2, joint analysis, etc.).

Sensitive data classification - Keeping sensitive data protected and abiding by strong research ethics are our highest priority. However, summary statistics from GWAS are not sensitive data. They cannot be used to identify participants. GWAS consortia producing summary statistics are composed of hundreds of thousands of individuals from potentially dozens of studies. The identifiability of a single individual from summary results of these mega-meta-analyses is impossible. Attempts to identify whether an individual is included in a cohort (per Homer et al. 2008) require access to individual-level genomic data and a sufficiently representative set of allele frequencies. While this is hypothetically possible, to then use that information to identify that participant’s individual level study data would be in violation of the data use agreement signed to gain access to the study dataset. Requiring investigators to affirm agreement with a statement regarding responsible use of the information in order to gain access to genomic summary statistics is a sufficient level of protection given the potential and hypothetical risk of identification (<https://www.genome.gov/pages/policyethics/genomicdata/aggdatareport.pdf>, see

finding 6). Gaining access to individual-level genomic data should be regulated and controlled, access to genomic summary results should not.

Summary statistics are also not subject to restrictions of use related to participant consents. Participants in studies may consent to use of their data only for a specific research question. But summary statistics are not participant data. Moreover, selected summary statistics – those deemed statistically significant or nearly so – are already being published. It is not clear why results that fall below some arbitrary statistical cutoff should be treated differently. No protections are gained by withholding the full set of genomic results.

Erin B. Ware, Survey Research Center, University of Michigan

Jessica D. Faul, Survey Research Center, University of Michigan

Colter Mitchell, Survey Research Center, University of Michigan

Kelly Bakulski, School of Public Health, University of Michigan

Daniel W. Belsky, Department of Population Health Sciences, Duke University School of Medicine

Arianna Gard, Department of Psychology, University of Michigan

Morgan Levine, Department of Pathology, Yale University

Lauren Schmitz, Population Studies Center, University of Michigan

Jennifer A. Smith, School of Public Health, University of Michigan

David R. Weir, Survey Research Center, University of Michigan

Wei Zhao, School of Public Health, University of Michigan

Homer N, et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet. 2008;4:e1000167. [[PMC free article](#)][[PubMed](#)]

Submission Date: 10/20/2017

Name: Steven Meyers

Name of Organization: GISAID Initiative (www.gisaid.org)

Type of Organization: Nonprofit Research Organization

Role: Institutional Official

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
 - Broad access to analysis of data is essential to complement rapid sharing and access to the original genetic data. In the experience of the GISAID Initiative, the benefits far outweigh any risks, when making influenza genetic data publicly accessible pre-publication for any analyses.
 - Having an effective licensing agreement in place, not only gives confidence to Data Providers in sharing data, but encourages users to reap benefits from analyses of the data. (*Shu et al* doi: 10.2807/1560-7917.ES.2017.22.13.30494)
 - Upholding any terms governing the access and use of genetic data (e.g. the GISAID sharing mechanism <http://gisaid.org/DAA>) can only be effective through the verification of user identity, and are not only to uphold the mechanism, but also to provide users with legal certainty on their rights to use the data, for example, for the publication of analyses, and development of medical interventions.
 - After nearly 10-years “GISAID has now developed a successful track-record in the field of influenza that may also serve as a useful blueprint for managing other diseases and global challenges requiring the international sharing of sensitive data.” (*Elbe et al* doi: 10.1002/gch2.1018)
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies

- The use of the term ‘unrestricted’ implies users may also use genomic summary results for purposes which may not be viewed as compliant with scientific etiquette, e.g. using the original data in a manner that fails to acknowledge the Provider, or making IPR claims to the detriment of the Provider, hence disincentivizing rapid sharing with the public.
- The use of the term ‘controlled-access’ must be defined whether it is:
 - a) controlled access, as in making access to the data available to the public with user identification as a controlling mechanism, or
 - b) controlled-access, as in making the data available to a consortium or selected category of individuals and institutions.
- GISAID strongly opposes the latter option, which bars the public’s immediate access to the data. The GISAID model demonstrates that its Open Access policy (re3data.org/DataCite doi.org/10.17616/R3Q59F) accomplishes both, with a measure of access control through positive user identity. This ensures compliance with the terms of use and provides incentive to Providers to share data.

Additional Comment (attachment):



RESEARCH ARTICLE

Global Health

Data, disease and diplomacy: GISAID's innovative contribution to global health

Stefan Elbe¹ and Gemma Buckland-Merrett²¹Centre for Global Health Policy, School of Global Studies, University of Sussex, Brighton BN1 9SN, UK²Centre for Global Health Policy, University of Sussex, Brighton BN1 9SN, UK

Impact Statement: The rapid spread of lethal infectious diseases is a global challenge potentially affecting any person around the world. To protect populations against such deadly outbreaks, it is critical that scientists and governments rapidly share information about the pathogens causing them. Without access to such information, it will be very difficult to properly assess the risk posed to global health, to develop new medical countermeasures, and to mount a commensurate international response. However, recent outbreaks suggest several impediments to the rapid sharing of virus data. Scientist may wish to withhold data until their scholarly studies are published; governments are fearful about the repercussions of being associated with a major new outbreak, and it remains challenging to fund global public goods like an international database to host such data. Through the first study of the Global Initiative on Sharing All Influenza Data (GISAID), this article shows how it is possible to encourage the greater international sharing of such data through the careful design of new sharing mechanisms. GISAID has now developed a successful track-record in the field of influenza that may also serve as a useful blueprint for managing other diseases and global challenges requiring the international sharing of sensitive data.

Keywords

data-sharing, GISAID, global health, influenza, pandemic preparedness, public-private partnerships, virus.

Correspondence

Stefan Elbe, Centre for Global Health Policy, School of Global Studies, University of Sussex, Brighton BN1 9SN UK.
Tel. 00441273678724; Fax: +44-1273-673563;
E-mail: s.elbe@sussex.ac.uk

Received: 12 April 2016;
Revised: 06 October 2016;
Accepted: 03 November 2016

doi: 10.1002/gch2.1018

Abstract

The international sharing of virus data is critical for protecting populations against lethal infectious disease outbreaks. Scientists must rapidly share information to assess the nature of the threat and develop new medical countermeasures. Governments need the data to trace the extent of the outbreak, initiate public health responses, and coordinate access to medicines and vaccines. Recent outbreaks suggest, however, that the sharing of such data cannot be taken for granted – making the timely international exchange of virus data a vital global challenge. This article undertakes the first analysis of the Global Initiative on Sharing All Influenza Data as an innovative policy effort to promote the international sharing of genetic and associated influenza virus data. Based on more than 20 semi-structured interviews conducted with key informants in the international community, coupled with analysis of a wide range of primary and secondary sources, the article finds that the Global Initiative on Sharing All Influenza Data contributes to global health in at least five ways: (1) collating the most complete repository of high-quality influenza data in the world; (2) facilitating the rapid sharing of potentially pandemic virus information during recent outbreaks; (3) supporting the World Health Organization's biannual seasonal flu vaccine strain selection process; (4) developing informal mechanisms for conflict resolution around the sharing of virus data; and (5) building greater trust with several countries key to global pandemic preparedness.

Introduction

The rapid spread of lethal infectious diseases is a global challenge potentially affecting any person living around the world. Already on multiple occasions in the 21st century, a deadly new infectious disease emerged suddenly and then quickly

spread through the dense network of international circulations that make up our globalized world – from HIV/AIDS and SARS, through pandemic flu and MERS, to recent experiences with Ebola and now the Zika virus.¹ Sharing

¹“Benefits of sharing”. *Nature*. 11 February 2016, p. 129.

Impact Box:

What challenges does the study address?

The timely international sharing of virus data is critical for protecting populations against lethal infectious disease outbreak. Without access to such information, it is very difficult to properly assess the risk posed to global health, to develop new diagnostics, medicines and vaccines, and to mount a commensurate international response. However, experiences with recent outbreaks suggest that there are three challenges when it comes to sharing virus data. First, scientists may hesitate to share data on lethal viruses because they are concerned about other researchers then using this data to publish scholarly articles more quickly than they can do themselves – meaning their scientific contribution is not properly acknowledged and recognized. Second, governments might interfere with the international exchange of information because of concerns about the negative economic ramifications of being identified as the source country of an international outbreak. They may also wish to retain ownership over any potential intellectual property associated with the data and – particularly for low- and middle-income countries – will be keen to ensure that they can secure access to new vaccines or medicines subsequently developed on the basis of that cooperation. Finally, there is also a more practical public goods challenge in terms of who will actually provide the funding and material infrastructure for hosting such virus data.

What is new about the research?

This research presents the first study of a new mechanism for encouraging the international sharing of virus data that has been created in the field of influenza. Initially spurred by the global threat posed by human infections with highly pathogenic avian influenza (H5N1), the Global Initiative on Sharing All Influenza Data (GISAID) was launched in 2008 as a new mechanism for incentivizing and promoting the international sharing of virus data.

What are the implications of the research?

The research shows how it is possible to overcome some of the challenges associated with the international sharing of virus data through the skillful design of new sharing mechanisms that are sensitive to the needs of stakeholders. Already, this important sharing mechanism has developed a successful track-record in the field of influenza and may also serve as a useful blueprint for other diseases and global challenges that depend on the international sharing of sensitive data. The research further shows how philanthropic actors can play an important role in bringing about novel global health initiatives and how important it is to build trust in new global health initiatives. Finally, the research also illustrates how innovative solutions to global challenges can be found when lessons are creatively applied from one issue area to another and that such cross-sectoral learning should be encouraged.

information on the viruses that cause such outbreaks is critical to protecting global health. Scientists need to rapidly share their information with other scientists around the world to understand the nature of the threat and to develop new medical countermeasures. Governments need the data to trace the extent of the outbreak, to coordinate public health interventions, and to ensure that populations have access to medicines and vaccines. All of this is particularly important in the case of flu, because of the comparatively rapid rate at which influenza viruses change, and the lurking spectre of a potentially devastating human pandemic. However, recent infectious disease outbreaks suggest that such international cooperation cannot be taken for granted and point to three impediments potentially hampering the timely sharing of such critical virus data.

First, scientists may hesitate to share data on lethal viruses because they are concerned about other researchers using this data to publish scholarly articles more quickly than they themselves are able to – meaning that their scientific contributions would not be properly acknowledged and recognized. Second, governments might also interfere with the international exchange of information because of concerns about the negative economic ramifications of being identified as the source country for an international outbreak. They may also wish to retain ownership over any intellectual property potentially residing in such data and – particularly for low- and middle-income countries – will wish to ensure that they can secure access to new vaccines or medicines developed on the basis of that cooperation. Finally, even where these two challenges can be overcome, there is still a much more practical obstacle in

terms of who will actually provide the international leadership, legitimacy, coordination, and funding needed for sustaining the material infrastructures essential for collecting, curating, and distributing such virus data. The international sharing of virus data may be critical to global health, but it is also enveloped in a Gordian knot of complex policy challenges.

The Global Initiative on Sharing All Influenza Data is an initiative aimed at untying that knot. Initially spurred by the global threat posed by human infections with highly pathogenic avian influenza (H5N1), GISAID introduced a new mechanism for incentivizing and promoting the international sharing of influenza virus data. GISAID's pivotal innovation consists of governing access to the data through a unique data access agreement extending a number of key "protections" and assurances to data contributors. In order to access the new database, users would first have to positively identify themselves through an initial registration and log-in process so that access to the data could be monitored. As part of that initial procedure, users would also agree to acknowledge those who submitted the data in their publications and to make best efforts to work collaboratively with data contributors on scientific publications. Users would further have to agree not to share the data with third parties outside the GISAID community and would also not seek to place any restrictions on the use of the data. Through extending such additional "protections," it was hoped that the GISAID mechanism could actively incentivize scientists and governments around the world to share influenza virus data in a timelier manner.

As GISAID marks its tenth anniversary, this article undertakes the first in-depth analysis of its wider contribution to global health. The primary source material for the analysis comes from more than 20 semistructured, background interviews conducted with key informants in the international community. Those stakeholders were drawn from the international scientific community, government institutions, pharmaceutical companies, research institutes, international organizations, and many of those involved in creating GISAID itself. Additional primary and secondary source material was identified in the form of policy papers, background papers, working papers, official documents, articles and books on virus sharing produced by science journalists, governments, think tanks, scientists, and international organizations. The article argues that GISAID is making at least five key contributions to global health: (1) collating the most complete repository of high-quality influenza data in the world; (2) facilitating the rapid sharing of potentially pandemic virus information during recent outbreaks; (3) supporting the World Health Organization's biannual seasonal flu vaccine strain selection process; (4) developing informal mechanisms for conflict resolution; and (5) building greater trust with several low-income and middle-income countries key to pandemic preparedness.

Obstacles to the international sharing of virus data

The international exchange of influenza virus data, including for viruses with human pandemic potential, is particularly important because influenza viruses evolve more rapidly than many other viruses.² Influenza, in other words, is a fast moving target. A World Influenza Centre was thus established at the Medical Research Council's National Institute for Medical Research in London in 1948. The World Health Organization (WHO) subsequently assumed the coordinating role for influenza virological surveillance with the establishment of the Global Influenza Surveillance Network (GISN) in 1952 – renamed the Global Influenza Surveillance and Response System (GISRS) in 2011.³ The network gradually expanded over subsequent decades, with a small number of laboratories becoming designated core WHO Collaborating Centres and receiving specimens/viruses for analysis from currently approximately 140 WHO-designated National Influenza Centres located in more than 100 countries around the world.

²Interview with Philip Dormitzer, Vice President and CSO: Viral Vaccines, Pfizer Vaccine Research and Development. 24 March 2015.

³Interview with Alan Hay, former Director of the WHO Collaborating Centre (World Influenza Centre) in London (1993 to 2009); Co-Chair of Scientific Advisory Council and Scientific Liaison Officer of GISAID. 15 September 2014.

Much of this early influenza work was carried out on the basis of biological characteristics of the viruses. As genetic sequencing technology became more widely available over the past two decades, however, so too genetic sequence data became more central to the process.⁴ At the same time, and with laboratories now sequencing more and more influenza viruses, it was also becoming evident that in practice much of these data were *not* being shared and *not* being made public.⁵ That is because there are at least three obstacles to the timely international sharing of influenza virus data.

Science, publications and recognition

The intensely competitive nature of science is one reason why virus data may not be shared in a timely manner. In a context where the standing of scientists, and the research income they can generate, is heavily linked to their publications, citations, and scientific reputations, there is pressure to be the “first” to publish findings – especially about a lethal new virus. Several interviewees expressed the view that scientists are concerned that sharing such information in an open and timely manner might enable others to publish findings with their data more quickly than they themselves could – meaning that their scientific contribution in discovering and analysing a new virus would not be properly credited and acknowledged (Longo and Drazen, 2016; Pearson, 2006: 963). Indeed, a number of interviewees expressed concerns about other researchers who just “crunch” the data generated and made available by others, without contributing to the generation of such data themselves,⁶ as well as the importance of end users giving appropriate credit to originators.⁷ Scientists from low-income and middle-income countries have also complained that analyses from samples they shared in the past (because they lacked the powerful molecular research capacity of laboratories in high-income countries) have subsequently been presented at international meetings and conferences without proper advance notification, or without including those who had shared the samples in the authorship arrangements (Sedyaningsih et al., 2008). Historically, some scientists have therefore decided to share such virus information in public databases only *after* their scientific papers had

⁴Interview with Alan Hay.

⁵Interview with John McCauley, WHO Collaborating Centre for Reference and Research on Influenza, Crick Worldwide Influenza Centre. 9 February 2015.

⁶Interview with Gwenaëlle Dauphin, Animal Health Service, Animal Production and Health Division, Food and Agriculture Organization of the United Nations, 9 October 2015; Interview with Yuelong Shu, Director of WHO Collaborating Centre for Reference and Research on Influenza, Beijing. 4 June 2014; Interview with John McCauley.

⁷Interview with Ian Brown, Director of EU/OIE/FAO International Reference Laboratory for Avian Influenza and Newcastle Disease. 16 September 2014; Interview with Alan Hay.

been published – leading to delays in the international sharing of data to the potential detriment of global health.⁸

This obstacle to the timely sharing of influenza virus data became all the more disconcerting when, in 2003, lethal human infections with highly pathogenic avian influenza H5N1 viruses reemerged in Hong Kong. Those human deaths raised the specter of a potentially much more devastating pandemic, with the human mortality rate of the virus reported by WHO at around 60% (WHO, 2016). With so much concern, fear, and attention now centring on the lethal H5N1 virus, the possible reputational benefits to scientists from being the first to publish analyses of the viruses were all the greater. It is precisely during such high-profile times that everyone wants to get there “first”.⁹ At the same time, the scientists at the forefront of such new outbreaks also suddenly become extremely busy, as their laboratories have to go into overdrive and often do not have enough resources to meet this surge in demand, whilst the priority in these circumstances is to generate science and evidence in support of control programmes.¹⁰ With an increased workload, scientists would have even less time than usual to write up their findings, meaning, there was a real risk that information critical to global health would not be shared quickly. Even though the need for sharing data is arguably much greater in the context of the threat of a human pandemic, so too are the obstacles to achieving such international sharing of virus data in practice.

Governments, trade and access to medicines

Governments may have additional reasons for not wanting data about lethal viruses to be shared internationally. They could be concerned with the negative economic ramifications of being identified as the county at the source of a lethal new outbreak,¹¹ and do not wish to be seen as the country that “caused” a devastating human pandemic.¹² There can also be intellectual property considerations surrounding such samples and information, which could be critical to the commercial development of new diagnostics, vaccines, and medicines.¹³

Indeed, low-income and middle-income countries in particular will be concerned to ensure access to any new medicines and vaccines produced with the help of such samples and data – as new medical countermeasures may later turn out to be too costly or only available in insufficient quantities.¹⁴

Many of these issues came to a fore in 2006 when, amidst intense concern about a possible flu pandemic, media reports surfaced about critical H5N1 sequence data not being made freely accessible to all countries – raising issues around fairness in accessing such data (Brown, 2006). It also emerged that new pandemic vaccines were being developed on the basis of biological samples initially obtained from affected countries in southeast Asia, but that originating countries were not consulted over the subsequent movement and sharing of such viruses with third parties (especially with industry) – raising additional concerns about the transparency of the GISN sharing mechanism (Sedyaningsih et al., 2008). Later, it transpired that once new H5N1 vaccines had been developed, they were not economically viable for many of those affected countries that had initially shared samples.¹⁵ Confronted with the prospect of having to ride out a flu pandemic without access to the same medical countermeasures available to many high-income countries, governments of affected countries began to openly question whom the sharing of virus samples through the GISN actually benefitted.

All these issues culminated in a lengthy and acrimonious diplomatic dispute over international virus sharing. The dispute was triggered when Indonesia (at the time experiencing the highest number of human cases of H5N1 infection) unexpectedly decided to stop sharing “its” virus samples from 20 December 2006 – marking the start of protracted, high-level diplomatic negotiations surrounding the equity, fairness, and transparency of influenza virus sharing (Supari, 2008: 24). In terms of global health security, Indonesia’s decision was regarded as a potential disaster in the sense that WHO now did not have a complete picture of how H5N1 was spreading and evolving¹⁶; but it also exposed the deeper political sensitivities around the international sharing of virus samples that would need to be addressed. As genetic sequence data of influenza viruses was becoming increasingly central to pandemic preparedness efforts, similar sensitivities emerged around such data as well.

⁸Interview with John McCauley. Interview with Nancy Cox, former Director of the Influenza Division, Centers for Disease Control and Prevention, Atlanta. 17 September 2014.

⁹Interview with Alan Hay.

¹⁰Interview with Ian Brown, Director of EU/OIE/FAO International Reference Laboratory for Avian Influenza and Newcastle Disease. 16 September 2014.

¹¹Interview with Ilaria Capua, former Head of the Virology Department at Istituto Zooprofilattico Sperimentale delle Venezie, Padova. 22 June 2015.

¹²Interview with Nancy Cox.

¹³Interview with Philip Dormitzer; Interview with Ron Fouchier, Professor in Molecular Virology, Department of Viroscience, Erasmus MC Rotterdam. 15 September 2014. See also Mary Quick. ‘Non-WHO global initiative on sharing avian influenza data’. *The Lancet Infectious Diseases* 6(10): 621. October 2006.

¹⁴Pearson, ‘Plan to pool bird-flu data takes off’, p. 963.

¹⁵‘Statement by the Minister of Health of the Republic of Indonesia H. E. DR. Dr. Siti Fadilah Supari’. November 2007. Available at: http://www.ip-watch.org/files/Indonesia_statement_WHO.pdf. [Accessed 11 January 2016].

¹⁶Interview with Robert Webster, Emeritus faculty, Department of Infectious Diseases St. Jude Children’s Research Hospital, Memphis. 10 July 2015; see also Richard Holbrooke and Laurie Garrett, ‘Sovereignty’ That Risks Global Health’. *Washington Post*. 10 August 2008.

Practical challenges: funding databases sustainably

There is also a much more practical obstacle to the sharing of virus data. Somebody needs to provide the international leadership, legitimacy, coordination, and funding necessary for maintaining the material infrastructures central to collecting, curating, and distributing such data. A database will require both a physical infrastructure, as well as scientific oversight to design the database, curate the information, liaise with laboratories around the world, and so forth. Yet several interviewees described just how difficult it is to secure funding for databases. They are not seen as glamorous or important as research projects,¹⁷ or indeed as an essential part of scientific intercourse that are central to advancement of scientific research.¹⁸ Unlike funding for research projects, moreover, databases also require continuous and even indefinite funding commitments, which many funders can be loath to agree to.¹⁹

Nevertheless, such an influenza database was initially created in the USA, by the Los Alamos National Laboratories in New Mexico, with funding from the US Centers for Disease Prevention and Control (CDC).²⁰ The database was created around the time of the first outbreak of lethal human infections with H5N1 in Hong Kong in 1997 and worked well for some years.²¹ However, when H5N1 began to spread internationally in 2004, it became clear that a number of affected countries were very sensitive about sharing H5N1 virus information and did not want WHO to release the data to others without their permission (Roos, 2006). As a way of acknowledging and addressing such concerns, those overseeing the Los Alamos database decided to create a separate, password-protected compartment from the seasonal flu database that would only be open to those working on sequencing H5 viruses.²²

Creating this private H5 compartment may have been a practical solution to the diplomatic sensitivities and tensions that were rapidly surfacing, but it also created new problems. Some scientists expressed frustration that without access to this compartment, they could not properly analyse how the viruses they were isolating related to the other viruses circulating internationally.²³ There was also a perception that the private compartment at Los Alamos was somewhat of a “club”, where only a limited group of scientists enjoyed access.²⁴ One scientist to

vocally draw public attention to this problem was the Italian veterinary scientist Ilaria Capua. In February 2006, her laboratory received a sample of an avian influenza virus infecting birds in Nigeria (Zamiska, 2006a). There was great scientific interest in the new sample because it was the first to come from Africa. Capua was thus offered access to the private compartment in return for sharing her findings with the Los Alamos database.²⁵ Capua, however, declined the offer and instead deposited the genetic sequence information in the public domain archive Genbank. Her decision to take a stand against that system, to challenge the status quo, and to invite other leading scientists to join her in pushing for a change of approach, attracted much scientific attention and media coverage at the time.²⁶ With unfavourable media attention surrounding the Los Alamos database, combined with the lack of a sustainable funding mechanism, as well as decreasing support from the US CDC, there was now a need to find a new home where such influenza virus data could be shared.²⁷

A new Global Initiative on Sharing All Influenza Data (GISAID)

The idea for an improved way of promoting the international sharing of influenza virus sequences was initially discussed in 2006, with the call for a new global initiative on sharing avian flu data (Zamiska, 2006b). Eventually, becoming the Global Initiative on Sharing *All* Influenza Data, GISAID's genesis was closely associated with Peter Bogner – a studio executive with a background in creating and licensing media content,²⁸ and in philanthropic behind-the-scenes work for organizations such as the United Nations and UNICEF.²⁹ Bogner provided the lion's share of funding for setting up GISAID (a low-mid seven figure sum) and was key to the development of the licensing mechanism that defines the GISAID data sharing policy.³⁰ As chief executive of its management board, he remains closely involved in the initiative to this day.³¹

¹⁷Interview with Ilaria Capua.

¹⁸Interview with Nancy Cox.

¹⁹Interview with Masato Tashiro, former Director, WHO Collaborating Centre for Reference and Research on Influenza, Japan, 5 June 2014; Interview with Catherine Macken.

²⁰Rock'n Rollouts'. *Multichannel New International*. November/December 1995; Hans-Juergen Jakobs, 'Der V-Faktor', *Sueddeutsche Zeitung*, 10 January 2005.

²¹Peter Bogner was Co-Chair of the UNICEF Entertainment Support Committee. Letter from Horst Cerni. 2. September 1988

²²Email correspondence with GISAID Initiative (Freunde von GISAID e.V.). 18. January 2016; Robin McDowell, 'Indonesia agrees to hand bird flu information to new online database'. *The Associated Press*. 16 May 2008.

²³Correspondence for the Meeting between German Government with GISAID. Bundesministerium fuer Ernaehrung und Landwirtschaft. 18. March 2015.

²⁴Interview with an expert in computational biology. 23 February 2015; Interview with a European expert in research and biotechnologies issues, formerly working in the European Commission. 19 February 2015.

²⁵Interview with Alan Hay.

²⁶Interview with Catherine Macken, Bioinformatics Institute of the University of Auckland, New Zealand. 17. December 2015.

²⁷Interview with Nancy Cox.

²⁸Interview with Ilaria Capua.

²⁹Interview with Nancy Cox.

³⁰Interview with Ron Fouchier.

³¹Interview with Ilaria Capua. See also Steven Salzberg, 'The Contents of the Syringe'. *Nature* 454. 10 July 2008, p. 161.

Remarkably, Bogner had no background in influenza – or even global health more broadly – prior to his involvement in GISAID, although he did have extensive experience with media and global affairs more generally. In 2006, a number of events would begin to draw him more closely into the world of influenza.

Bogner's interest in pandemic preparedness was originally piqued during a breakfast with Michael Chertoff (then United States Secretary of Homeland Security) and a small group of executives at the World Economic Forum in Davos in January 2006. Not long thereafter, Bogner was also asked by the New York Office of the United Nations Secretary General to use his extensive media contacts to look into unfavourable media claims alleging that WHO was operating a “secret” database in Los Alamos for genetic sequences of highly pathogenic avian influenza (HPAI) viruses and first liaised with Dr. Margaret Chan, who would subsequently become Director General of the WHO.³² In April 2006, Bogner then attended a scientific influenza meeting in Cambridge where he met Nancy Cox, who that year became head of the influential influenza division at US CDC. Sharing a ride with Bogner back to London, Cox had an opportunity to elaborate upon the challenges surrounding pandemic preparedness and the importance of international influenza sample and data sharing (Zamiska, 2006b).

As Bogner became more understanding of the complexities involved, he also learned from Cox of Indonesia's dilemma surrounding a particularly disconcerting H5N1 virus outbreak in Karo, Sumatra – where limited human to human transmission could not be immediately ruled out. Given his political connections in Indonesia, Bogner was able to secure a meeting to speak directly with the Indonesian Minister of Health Siti Supari, who was leading Indonesia's approach to these issues. Eventually Bogner was even able to persuade Supari to share the sequences of the Karo cluster, resulting in their deposit in Genbank.³³ It was an unexpected diplomatic breakthrough, and a decision quickly met with reciprocity by the US CDC, when Cox not long thereafter announced that they too would be making influenza data publically accessible (Quick, 2006: 621). By this time, Bogner had become deeply immersed in the issues around the international sharing of influenza virus data.³⁴

The creation of GISAID moved a step closer when the influential scientific journal *Nature* published a prominent letter signed by more than 70 leading scientists (including seven

Nobel laureates) in August 2006. The letter – coauthored by Peter Bogner, Ilaria Capua, Nancy Cox, and David Lipman – proposed the creation of a new global consortium that would foster international sharing of avian influenza isolates and data (Bogner et al., 2006). The letter, whose signatures included many researchers and officials from countries directly affected by H5N1, stated the intention for scientists participating in the consortium to share their sequence data, to analyse findings jointly, and to publish results collaboratively (ibid.). Although initially only members of the consortium would be able to access the data, as soon as possible following analysis and validation (and no longer than six months later), the data would then be deposited in publicly available databases that are part of the International Sequence Database Collaboration (e.g. EMBL, DDBJ, and Genbank) (ibid.).

Bogner was prominently listed as first author of the *Nature* letter and coined the acronym GISAID.³⁵ However, along with a number of other influenza scientists, he was also aware that notwithstanding its good intentions, the brief letter still lacked much practical detail, and that the core issues of transparency and equity of data sharing would likely remain unresolved if data archives with anonymous access to data (like Genbank) were used. Ultimately, they felt, the sharing of such data would only work if any rights to the data that may exist would remain untouched through the process of sharing.³⁶ Moving things forward in practice would thus necessitate much more extensive consultation with a range of different stakeholders and mediating the development of a new system satisfying their various needs and concerns. At this point, Bogner set out to use his considerable knowledge of media and licensing issues, along with a pool of legal experts in intellectual property, government lawyers, as well as the help of key influenza experts, to develop a new mechanism that would permit the sharing of data without delay in a publicly accessible and free database, yet to be developed.³⁷ Over the next 18 months, those efforts would focus on three key areas: (1) developing and negotiating with Members States the legal terms for a new database access agreement; (2) the technical design of a new influenza database; and (3) agreeing the initiative's governance structure.

An innovative approach to data sharing: the GISAID data access agreement

How could a new system better incentivize the international sharing of virus data? One possible way forward would be to

³²Email Correspondence between Margaret Chan and Peter Bogner. 17 March 2006.

³³CIDRAP, 'Indonesia, FAO, OIE pledge to publish H5N1 data'. 3 August 2006. Available at: <http://www.cidrap.umn.edu/news-perspective/2006/08/indonesia-fao-oie-pledge-publish-h5n1-data>. [Accessed 11 January 2016]; Email correspondence between Peter Bogner and Nancy Cox. 3 August 2006; Email correspondence between John Sulston and Peter Bogner. 5 August 2006.

³⁴McDowell, “Indonesia agrees to hand bird flu information”.

³⁵Interview with Peter Bogner, Chairman of GISAID (Freunde von GISAID e.V.). 28 November 2013.

³⁶Ibid.

³⁷Ibid.

try and provide data contributors with additional protections and assurances about how their data would be used. This could be performed through the careful legal design of a new data agreement governing the submitters' deposits of, and users' access to, such influenza virus data. From late 2006 onwards, Bogner thus engaged a number of former colleagues and lawyers who had worked with him on intellectual property issues during his broadcasting days to help realize a new data licensing agreement.³⁸ Additional scientific input to this agreement, particularly in terms of providing the scientific language for the sharing agreement and helping to define the data, came from the Scientific Advisory Council (SAC). Constituted early on by GISAID, the SAC was initially chaired by Nancy Cox from the US CDC and is composed of fellow WHO Collaborating Centre directors and FAO/OIE Reference Laboratory counterparts, as well as established researchers in the fields of epidemiology, human virology, veterinary virology, and bioinformatics. The SAC is now co-chaired by Nancy Cox and John McCauley of the Crick Institute.

Developing this new data access agreement (DAA) would require Bogner to travel extensively around the world in an effort to forge an international consensus on a novel sharing mechanism enabling public and animal health authorities to continue their surveillance work, that ensured manufacturers of medical countermeasures could continue their work on developing vaccines, antivirals, or diagnostic kits, and that also provided a transparent mechanism for researchers who had the publication of their manuscripts as their main focus.³⁹ GISAID's resulting DAA, which came into force in May 2008, retained the principle of a publicly accessible database – meaning that any natural person (whether scientist or not) could obtain credentials to access data in GISAID, predicated upon a one-time positive verification of the individual's identity, and agreement to the terms of the DAA, which license the use of data in GISAID.⁴⁰ This process of positively identifying the contributors and users of data differs from the anonymous access afforded to public domain archives (like Genbank), but provides GISAID with the mechanism for enforcement, and makes users adhere to the rules set forth in the DAA. Further benefits of this system are that it makes it easier for scientists to discover and properly acknowledge those who contributed the data and to also assist with any biosecurity considerations that could potentially arise around the use of some such data.⁴¹

³⁸Ibid.

³⁹McDowell, 'Indonesia agrees to hand bird flu information'.

⁴⁰GISAID EpiFlu™ Database Access Agreement © 2008–2016. Freunde von GISAID e.V. Available at: <http://gisaid.org/DAA>. [Accessed 23 March 2016].

⁴¹Email correspondence with GISAID Initiative (Freunde von GISAID e.V.). 26. September 2016. See also Lawrence O. Gostin, Alexandra Phelan, Michael A. Stoto, John D. Kraemer and Srinath Reddy. 'Virus sharing, genetic sequencing, and global health security'. *Science*. 345(6202): 1295–1296.

The core provisions of the DAA thus include that users: (1) will share their own data and allow other users to access it; (2) that they will *not* share or distribute data submitted directly to the GISAID sharing mechanism to other non-GISAID servers or to individuals/institutions who are not registered GISAID users; (3) that they will credit the use of others' data in publications; (4) that they will make best efforts to collaborate with the originating laboratory and involve them in analyses and further research involving the data; (5) that they will analyse findings jointly; and (6) that they will maintain common access to technology derived from the data so that it can be used not only for research but also for the development of medical interventions such as diagnostics, vaccines, or antivirals. According to the agreement, GISAID users thus have the right to develop a commercial product on the basis of data obtained through GISAID, but they may not impose any terms on the data itself (which remains the sole property of the contributor), and they must also seek to collaborate with the data contributors.⁴²

Most crucially of all, and notwithstanding its status as a publically accessible database, GISAID would therefore *not* fall under the legal definition of "Public Domain", because the GISAID license respects the ownership of data submissions by explicitly not permitting the removal – or waiving – of any potential pre-existing "rights" to the data; to the extent that such rights might exist around the data, they would not be affected by virtue of them having been submitted to GISAID.⁴³ The unique sharing mechanisms thus ensure that inherent rights (such as intellectual property rights) are not forfeit when sharing data.⁴⁴ In many ways, the successful development of this DAA offering additional legal protections and clarity would mark the key to GISAID's new virus sharing mechanism.

EpiFlu™: creating a new influenza database

In parallel to the legal access agreement, it would also be necessary to develop the actual database itself – especially with the closing of the existing database at Los Alamos. Decisions would have to be made about where to physically locate the new database, how to design its structure and features, as well as making practical arrangements for its day-to-day running. In February 2007, it was announced that the GISAID initiative would collaborate with a Swiss consortium consisting of the Swiss Institute of Bioinformatics (SIB), and the Swiss bioinformatics company SmartGene. SmartGene would provide

⁴²Email correspondence with GISAID Initiative (Freunde von GISAID e.V.). 26 September 2016.

⁴³Interview with Peter Bogner.

⁴⁴Statement of the Federal Republic of Germany. 'On Substantive Issues and Concerns Regarding the PIP Framework and ITS Implementation'. Special Session of the PIP Advisory Group, 13 October 2015.

secure storage and analysis of the influenza data, whilst SIB would complement genetic sequence information with high-quality protein annotation provided by the team of lead scientist Amos Bairoch, who was well known for having developed the Swiss-Prot protein knowledge database.⁴⁵ Those initial arrangements for a new database would later become embroiled in complex legal disputes around contractual obligations and the flow of public funds from the USA, via the World Health Organization.⁴⁶ This resulted in GISAID establishing a new EpiFlu™ database in Germany in 2009, following the German government's proposal to ensure GISAID's continuation by acting as its new official host.⁴⁷

To design the new database, GISAID used the same technical staff (today's Database Technical Group) composed of experts who held daily responsibilities for the sequencing activities at leading institutions such as the US' CDC, the National Institute for Medical Research in London, as well as the WHO Collaborating Centres in Beijing, Melbourne, and Tokyo, or who had worked on the design of the now defunct Los Alamos Influenza Database (Schnirring, 2009). The new EpiFlu™ database in Germany was then developed by the Max Planck Institut for Informatics (in Saarbruecken) in consultation with the wider scientific community.⁴⁸ Because the data that had been uploaded into the original EpiFlu™ database still belonged to those who supplied it (rather than to SIB), the data could then be migrated to the new database in Germany following the consent of the original contributors (Greenemeier, 2009).

With the move to Germany, responsibility for hosting the EpiFlu™ database and GISAID platform would henceforth rest with the German government.⁴⁹ Four German institutions in particular would play central roles: (1) the Federal Ministry of Food and Agriculture representing Germany; (2) the Friedrich Loeffler Institute (Germany's Federal Research Institute for Animal Health) responsible for data quality; (3) the Federal Office for Agriculture and Food for the technical hosting of the database; and (4) the Max Planck-Institute for Informatics, which would develop a new software application. In 2011, the

Federal Republic of Germany and GISAID also announced that the German government would be the long-term host of the EpiFlu™ database and GISAID platform, which it continues to do to this day.⁵⁰ Ensuring the proper design, implementation, and sustainability of the new EpiFlu™ database thus marked a second key dimension in GISAID's data sharing mechanism.

Developing a governance structure

Establishing an appropriate governance structure formed the final element. Given the scientific and political sensitivities surrounding influenza virus data, a proper governance structure would be vital to ensuring the legitimacy, scientific credibility, independence, and sustainability of the new initiative. Indeed, without such a structure, it would be difficult to build the requisite level of trust amongst scientists and governments necessary for them to agree to share the data with the new initiative. GISAID's governance structure would thus come to comprise of three independently operating bodies: (1) a board of trustees charged with securing the independence of GISAID from political or commercial interests; (2) the Scientific Advisory Council, providing scientific inputs and oversight of the initiative; and (3) the Database Technical Group, offering expertise in developing the database.⁵¹

In the end, then, it took a good year and a half to move from the initial aspiration for a new global consortium expressed in the 2006 *Nature* letter, to finalizing all of the careful legal, practical, and governance arrangements needed for launching a new virus data sharing platform. Getting the GISAID sharing mechanism off the ground ended up being much more than just a technical challenge of developing a new influenza database. Although the data and database remain central to the enterprise, GISAID represents a much wider international initiative comprising the EpiFlu™ database alongside its innovative sharing mechanism (enshrined in its database access agreement), as well as its wider governance structure. All three elements are critical to achieving its aim of actively promoting the timely international sharing of all influenza virus data. Once these elements were in place, GISAID could officially be launched in Geneva on the occasion of the 61st World Health Assembly in May 2008.⁵²

The initiative's activities in different countries (especially China, Indonesia, and the USA) were later streamlined by

⁴⁵'Swiss Consortium to Manage GISAID Database.' GISAID/Swiss Institute of Bioinformatics/SmartGene Press Release. 19 February 2009; Email correspondence between Amos Bairoch and Peter Bogner. 27 September 2006.

⁴⁶Robin McDowell, 'Influenza Scientists, WHO face off in virus row'. The Associated Press. 3 October 2008.

⁴⁷Aigner unterstützt Ansiedlung einer internationalen Influenza-Datenbank in Bonn'. German Federal Ministry of Food and Agriculture. 16 October 2009. Available at: <http://www.bmel.de/SharedDocs/Pressemitteilungen/2009/249-Ai-Influenza-Datenbank%20Bonn.html> [Accessed 23 March 2016]

⁴⁸'GISAID Launches Second Influenza Database'. GISAID Foundation Press Release. 14 September 2009. Available at: <http://www.prlog.org/10340963-gisaid-launches-second-influenza-database.html>. [Accessed 13 January 2016].

⁴⁹CIDRAP, 'German government to host flu database', 20 October 2009 (<http://www.cidrap.umn.edu/news-perspective/2009/10/german-government-host-flu-database>) [Accessed 07 January 2016].

⁵⁰German Federal Ministry of Food and Agriculture. (http://www.bmel.de/EN/Ministry/Research-Innovation/_Texte/Influenza-Datenbank-EpiFlu.html) [Accessed 07 January 2016].

⁵¹GISAID, "Publicly accessible EpiFlu database featuring the world's most complete collection of Influenza data" (http://www.ble.de/SharedDocs/Downloads/06_Dienstleistungen/07_DienstleistungszentrumIT/100420_GISAID-Flyer.pdf?__blob=publicationFile) [Accessed 30 November 2015].

⁵²GISAID, "Publicly accessible EpiFlu™ 2.0 Database featuring the world's most current and complete collection of Influenza Data". Available at: http://www.influenza.spb.ru/files/GISAID_Flyer.pdf [Accessed 08 January 2016].

formalizing them into a registered nonprofit association in Germany operated exclusively for charitable, scientific, and educational purposes – called “Freunde von GISAID e.V. [Friends of GISAID]”⁵³ In 2013, the German government also reaffirmed its long-term commitment to host the GISAID platform and the EpiFlu™ database, ensuring its sustainability.⁵⁴ Wider operations, including user management, bilateral consultations with member states and dialogue with international organizations, or scientific matters remain GISAID’s responsibility.⁵⁵ GISAID similarly handles registrations and technical support questions.⁵⁶ GISAID today also has other public-private partnerships – with the US Centers for Disease Control & Prevention and Singapore’s Agency for Science, Technology & Research – which contribute to the development of technology and the educational programmes of the initiative.⁵⁷

GISAID’s contribution to global health

Has GISAID been successful in meeting its aims and objectives? As the initiative marks its 10th anniversary, there is substantial evidence of a sustained track-record of successfully facilitating the international sharing of influenza virus data. Indeed, five contributions of GISAID to global health stand out: (1) collating the most comprehensive repository of influenza genetic sequences, as well as associated clinical and epidemiological metadata; (2) facilitating the rapid sharing of potentially pandemic virus data during recent outbreaks; (3) supporting the WHO’s bi-annual influenza vaccine virus recommendation; (4) developing informal mechanisms for conflict resolution around the sharing of virus data; and (5) building greater trust with many low- and middle-income countries key to global pandemic preparedness.

⁵³GISAID, “Publically accessible EpiFlu database featuring the world’s most complete collection of Influenza data” (http://www.ble.de/SharedDocs/Downloads/06_Dienstleistungen/07_DienstleistungszentrumIT/100420_GISAID-Flyer.pdf?__blob=publicationFile) [Accessed 30 November 2015]; Correspondence with GISAID Initiative (Freunde von GISAID e.V.). 3 October 2016; Federal Ministry of Food and Agriculture. ‘BMEL unterstützt den Kampf gegen Influenzaviren’. Press Release. 5 November 2014. Available at: <http://www.bmel.de/SharedDocs/Pressemitteilungen/2014/277-KL-GISAID.html>. [Accessed 6 October 2016].

⁵⁴Correspondence with GISAID Initiative (Freunde von GISAID e.V.). 3 October 2016.

⁵⁵Cooperation Agreement between Freunde von GISAID e.V. [Friends of GISAID] and the Bundesministerium fuer Ernährung und Landwirtschaft [German Federal Ministry of Food and Agriculture]. 5 November 2014, p. 3.

⁵⁶Ibid.

⁵⁷Correspondence with GISAID Initiative (Freunde von GISAID e.V.). 3 October 2016.

Comprehensive international influenza virus data

Since its formal launch GISAID has rapidly built up an international user base comprising more than 6500 users.⁵⁸ Those users today span individuals at universities, research institutes and public health organizations, clinicians, animal health experts, bioinformaticians, epidemiologists, and members of industry from around the world.⁵⁹ The fact that GISAID is now widely used by the WHO – comprising the WHO Collaborating Centres for Influenza, the world’s National Influenza Centers, and others⁶⁰ – has helped GISAID collate the most complete repository of high-quality influenza data in the world.⁶¹ Influenza data curated by a combination of automatic and manual steps from more than 850 institutions are now held and governed by the GISAID database access agreement.⁶²

The EpiFlu™ database today contains the genetic sequences of more than 1,000 influenza viruses with Human Pandemic Potential (IVHPP).⁶³ The most recent human influenza sequences – human isolates of the subtypes H5N1, H6N1, H7N3, H7N7, H7N9, H9N2, H10N8, and H3N2 – are all contained in the database.⁶⁴ The database also holds sequences from other hosts – with (as of 2014) humans making up approximately 69% of data, avian species 19%, and other mammals like swine 10%.⁶⁵ Geographically, data is received from around the world, with approximately 36% of submissions coming from Asia, 29% from North America, and 22% from Europe, 5% from Oceania, 4% from Africa and

⁵⁸Email correspondence with GISAID Initiative (Freunde von GISAID e.V.). 26. September 2016.

⁵⁹PIP Framework Advisory Group, ‘Draft optimal characteristics of an influenza genetic sequence data sharing system under the pip framework’. 2015. Available at: http://www.who.int/influenza/pip/advisory_group/draft_twg_doc.pdf. [Accessed 3 December 2015].

⁶⁰WHO Global Influenza Surveillance and Response System (GISRS). Available at: http://www.influenzacentre.org/centre_GISRS.htm. [Accessed 5 March 2016].

⁶¹PIP Framework Advisory Group, ‘Draft optimal characteristics of an influenza genetic sequence data sharing system under the pip framework’. 2015. Available at: http://www.who.int/influenza/pip/advisory_group/draft_twg_doc.pdf. [Accessed 3 December 2015].

⁶²Email correspondence with GISAID Initiative (Freunde von GISAID e.V.). 26. September 2016.

⁶³The GISAID Initiative website (<http://gisaid.org>) [Accessed 11 January 2016]; PIP Framework Advisory Group, ‘Draft optimal characteristics of an influenza genetic sequence data sharing system under the pip framework’. 2015. Available at: http://www.who.int/influenza/pip/advisory_group/draft_twg_doc.pdf. [Accessed 3 December 2015].

⁶⁴GISAID, “EpiFlu™ database features and tools presentation” WHO-GISAID training workshop, St Petersburg 28–29 August, 2014, WHO-GISAID training workshop. Available at: <http://www.influenza.spb.ru/en/conferences/gisaid-2014-en> [Accessed 30 November 2015].

⁶⁵Ibid.

4% from South America (as of 2014).⁶⁶ At the end of 2015, the EpiFlu™ database broke through the symbolic threshold of holding 500,000 genetic sequences.⁶⁷ By October 2016, this number had risen to more than 650,000 genetic sequences.⁶⁸

Data directly submitted to the EpiFlu™ database are also regularly complemented with sequence data deposited into public domain archives that are part of the International Nucleotide Sequence Database Collaboration (covering DDBJ, EMBL-EBI, and NCBI). Of the 172,322 virus isolates held in the EpiFlu™ database (as of 4 October 2016), around 40% (65,915) were submitted to GISAID directly, and approximately 60% (106,407) were initially uploaded through the International Nucleotide Sequence Database Collaboration (INSDC).⁶⁹ The EpiFlu™ database thus holds both the sequence information submitted directly through the GISAID platform, as well as those submitted to INSDC databases. In light of GISAID's emphasis on hosting high-quality data, the latter data go through a process of further curation before being incorporated into the EpiFlu™ database. Given how rapidly influenza viruses change, moreover, the ability of the GISAID sharing mechanism to attract recent data is particularly significant. Of the data contained in the EpiFlu™ database from viruses collected solely over the past year, 81% were submitted directly to GISAID – and 93% for viruses collected over the last 6 months.⁷⁰

Beyond genetic sequence data, the EpiFlu™ database also stores and provides around 50 different fields of associated meta-data (of which most are searchable).⁷¹ Such metadata includes date of specimen collection, specimen source, date of virus harvest, antiviral susceptibility, and – for human samples – patient information such as age, gender, and health status.⁷² Along with a number of data analysis tools integrated into the GISAID platform, these additional data are seen by many researchers to be valuable features of the EpiFlu™ database, including for the purposes of surveillance and pandemic preparedness.⁷³ At the time of writing, plans were also underway to launch the next-generation of the database – EpiFlu™ 2.0 – with funding provided from 2013 to 2016 through the European Union's Research and Innovation funding programme.⁷⁴ According to

the directors of the core WHO Collaborating Centres for Influenza, GISAID has rapidly emerged as an essential resource and an “irreplaceable cornerstone for public and animal health in the global fight against influenza” upon which the influenza community now depends.⁷⁵

Global health security: encouraging rapid data sharing during outbreaks

GISAID has also demonstrated its ability to promote the rapid sharing of virus data during several key outbreaks of wider concern to global health security. Indonesia's 2008 decision to resume sharing H5N1 data through GISAID was hailed as a diplomatic triumph.⁷⁶ GISAID would again play a key role in the 2009 outbreak of pandemic H1N1. In April 2009, the first news of the novel H1N1 virus initially threatened to overwhelm the database, yet registrations were maintained around the clock to ensure everyone had access to the data and could share early detection findings with health authorities.⁷⁷ By April 25, the US CDC had uploaded the first full genome sequence of the new virus from initial US cases into GISAID, instantly giving the research community its first detailed look at novel H1N1.⁷⁸ That information was also used for developing new diagnostics for the virus,⁷⁹ as well as for subsequent attempts to develop new vaccines for pandemic H1N1.⁸⁰

More recently, GISAID was again used for the rapid sharing of data about the potentially pandemic H7N9 virus that caused human deaths in China and therefore also raised significant international concern. China reported the H7N9 outbreak to the WHO on 31 March 2013, just 6 weeks after the first known person fell ill.⁸¹ On the same day, it published the genomic sequences of viruses from the three human cases then identified on the database of GISAID, along with sharing the data and live virus with the WHO GISRS and other

⁶⁶Ibid.

⁶⁷The GISAID Initiative website (<http://gisaid.org>) [Accessed 7 January 2016].

⁶⁸Search of EpiFlu Database performed on 4 October 2016.

⁶⁹Search of EpiFlu Database performed on 4 October 2016.

⁷⁰Search of EpiFlu Database performed on 4 October 2016.

⁷¹Email correspondence with GISAID Initiative (Freunde von GISAID e.V.). 26. September 2016.

⁷²PIP Framework Advisory Group, 'Draft optimal characteristics of an influenza genetic sequence data sharing system under the pip framework'. 2015. Available at: http://www.who.int/influenza/pip/advisory_group/draft_twg_doc.pdf. [Accessed 3 December 2015].

⁷³Interview with animal health specialist. 16 September 2014.

⁷⁴European Commission PREDEMICS Consortium. Available at: http://cordis.europa.eu/project/rcn/101795_de.html [Accessed 23 March 2016]

⁷⁵Letter of the Directors of the World Health Organization Collaborating Centres for Influenza to the German Government. 18 September 2015.

⁷⁶Interview with Masato Tashiro.

⁷⁷CIDRAP News, 'Pandemic reveals strengths of new flu database' (2009). Available at: <http://www.cidrap.umn.edu/news-perspective/2009/06/pandemic-reveals-strengths-new-flu-database>. [Accessed 26 November 2015].

⁷⁸Ibid.

⁷⁹Viral gene sequence to assist update diagnostics for swine influenza A (H1N1). Geneva: World Health Organization. 25 April 2009.

⁸⁰'CDC Testing New H1N1 Vaccine Target' *Recombinomics Commentary* 3 August 2011. Available at: http://www.recombinomics.com/News/08031101/H1N1_Q226R_Vaccine.html. [Accessed 3 March 2016]; and 'WHO Testing New Pandemic H1N1 Vaccine Target'. *Recombinomics Commentary* 18 August 2011. Available at: http://www.recombinomics.com/News/08181101/H1N1_Vaccine_WHO.html. [Accessed 3 March 2016].

⁸¹ECDC, 'Human infection with a novel avian influenza A(H7N9) virus, China', 27 January 2014. Available at: <http://ecdc.europa.eu/en/publications/Publications/influenza-AH7N9-China-rapid-risk-assessment-27-January-2014.pdf>. [Accessed 11 January 2016].

laboratories.⁸² Using the data, the new virus genes could be synthesized in the USA in a matter of days, thus enabling the vaccine company Novartis to also rapidly develop a new vaccine.⁸³ Since that time, other companies have also used data in GISAID to develop new H7N9 vaccines.⁸⁴ GISAID thus plays a crucial role in the timely exchange of information integral to the selection of pre-pandemic vaccine viruses.⁸⁵

Strain selection for the seasonal flu vaccines

Beyond the specter of pandemic flu, EpiFlu™ also contains the most up-to-date collection of data on seasonal influenza viruses (Neher and Bedford, 2015). The EpiFlu™ database generally receives submissions of data from current/novel strains significantly quicker than data generated from retrospective studies (ibid.). The EpiFlu™ database is thus routinely used during the bi-annual process of selecting viruses that will form the basis for the seasonal flu vaccine in the Northern and Southern Hemispheres. Data on seasonal human influenza viruses for the biannual vaccine strain consultation meetings (VCM) are deposited by WHO Collaborating Centres within a time-frame of days to a few weeks of sequencing, depending on urgency and other circumstances.⁸⁶ Several of the WHO Collaborating Centres first used the database in September 2008 (following its earlier launch in May of that year) to make their recommendations for the Southern Hemisphere 2009 seasonal flu vaccine, and all of the centres subsequently used it in February 2009, to make recommendations for the Northern Hemisphere 2009–2010 vaccine.⁸⁷ Since that time, GISAID has been consistently relied upon by WHO Collaborating Centres for this selection process – giving the database simultaneous global health utility for pandemic *and* seasonal flu.⁸⁸

⁸²‘The fight against bird flu’ *Nature* (2013). Available at: <http://www.nature.com/news/the-fight-against-bird-flu-1.12850>. [Accessed 25 November 2015].

⁸³‘Vaccine News Daily, ‘Positive Phase I clinical trial results for Novartis H7N9 vaccine’. 19 November 2013. Available at: <http://vaccinenewsdaily.com/stories/510535576-positive-phase-i-clinical-trial-results-for-novartis-h7n9-vaccine> [Accessed 11 January 2016].

⁸⁴‘Medicago Produces VLP Vaccine Candidate for Emerging H7N9 Virus’. *Global Biodefense*. Available at: <http://globalbiodefense.com/2013/05/09/medicago-produces-vlp-vaccine-candidate-for-emerging-h7n9-virus/> [Accessed 3 March 2016]; ‘Vaxart develops H7N9 vaccine’. *Vaccines News Daily*. 28 June 2013. Available at: <http://vaccinenewsdaily.com/stories/510534785-vaxart-develops-h7n9-vaccine>. [Accessed 3 March 2016].

⁸⁵‘Letter of the Directors of the World Health Organization Collaborating Centres for Influenza to the German Government. 18 September 2015.

⁸⁶CIDRAP News, ‘Pandemic reveals strengths of new flu database’ (2009) Available at: <http://www.cidrap.umn.edu/news-perspective/2009/06/pandemic-reveals-strengths-new-flu-database>. [Accessed 26 November 2015].

⁸⁷Ibid.

⁸⁸Interview with Nancy Cox; ‘Virendatenbank GISAID – Global Initiative on Sharing All Influenza Data’. Max Planck Foundation. Available at: <http://www.maxplanckfoerderung.org/project/333/>. [Accessed 23 March 2016].

Resolving potential conflicts amongst stakeholders

Over the years, GISAID has also evolved a track record of successfully managing potential tensions and conflicts between stakeholders within the consortium. This became especially clear over the 2013 sharing of Chinese H7N9 virus data, which led to two sets of tensions. First, it emerged on 5 April 2013 that the vaccine company Novartis and the J. Craig Venter Institute were planning to use the sequences uploaded into GISAID to develop a new H7N9 vaccine with the support of the US CDC and with funding from the US government (Butler and Cyranoski, 2013). Because of time constraints and immense concern about the lethality of the virus, however, they initially proceeded without involving the Chinese researchers. They, in turn, felt that this move was not in the spirit of GISAID’s data access agreement, which required data users to make best efforts to collaborate with the originating laboratory responsible for obtaining the specimens. At this point, GISAID’s president stepped in and was able to mitigate the situation because of his close ties with the parties involved. Following discussions between the parties, the Chinese communicated to GISAID that they were satisfied that Novartis and its partners were engaging with China in a collaborative effort, and the vaccine development plans were able to proceed.⁸⁹ Bogner was later commended for his efforts in this matter by both the former Head of Research at Novartis, Philip Dormitzer, as well as the head of the Chinese National Influenza Centre, Yuelong Shu.⁹⁰

A second source of tension emerged around the same time when the Chinese scientists submitted their first major scientific paper on H7N9, including analyses of the sequences, to the prestigious *New England Journal of Medicine*. The Chinese researchers learned that they might be scooped, as a major analysis of the H7N9 virus was already due to come out in the journal *Eurosurveillance* on 10 April, with Masato Tashiro (then director of the WHO’s Collaborating Centre in Tokyo) as a co-author. Tashiro claims a draft of the paper was sent to the Chinese researchers along with an offer of co-authorship, which was declined.⁹¹ Bogner again played a key role behind the scenes in brokering a solution by effectively raising concerns of scientific etiquette that could be amicably resolved. Tashiro was asked to delay publication until after the Chinese research publication on 11 April, and their publication was published later – albeit still on the same day (Butler and Cyranoski, 2013). Although the episode confirmed the continuing tensions that exist around the international sharing of virus data, it also showed that such tensions could be constructively managed within the framework and spirit of GISAID.

⁸⁹Email correspondence with GISAID Initiative (Freunde von GISAID e.V.). 26. September 2016.

⁹⁰Ibid.

⁹¹Interview with Yuelong Shu; D. Butler and D. Cyranoski, ‘Flu papers spark row over credit for data’, *Nature*, 497 (2013), pp. 14–15.

Building trust internationally

Despite being a relative newcomer to the global health landscape, GISAID has already garnered a reputation for building trust and respect internationally.⁹² The fact that a substantial proportion of the influenza sequences deposited in GISAID's database originate from Asia and Africa is no coincidence.⁹³ GISAID gave many LMICs the feeling that their concerns matter and that users of GISAID are treated equally.⁹⁴ The list of countries that have submitted data thus includes laboratories located in Vietnam, Brazil, Argentina, Cambodia, Thailand, India, Chile, Kenya, and Morocco.⁹⁵ It also includes countries that are deemed, because of their geographical location and role in responding to past outbreaks, to be central to global pandemic preparedness – including Indonesia, Mexico, Egypt, and China (including Hong Kong).⁹⁶ Overall, 201 countries and territories around the world participate in GISAID.⁹⁷

This trust from many low-income and middle-income countries is further nurtured through international workshops instructing researchers from around the world how to work and analyse data generated from viruses isolated in their region. To this end, GISAID has partnered with a number of other organizations such as the WHO's GISRS, the Antiviral Group of the International Society for Influenza and Other Respiratory Viruses Diseases (ISIRV), the PREDEMICS Consortium, and the Tan Tock Seng Hospital to host workshops in Africa, Asia, Russia, Europe, and the USA.⁹⁸ Those workshops help to build further trust (as well as capacity) around the international sharing of influenza virus information amongst researchers from low-income and middle-income countries.

GISAID has also enjoyed other forms of international recognition. A meeting of the Association of South East Asian Nations (ASEAN) recognized GISAID in 2009 for encouraging

the sharing of influenza genetic data.⁹⁹ Its role in moving towards greater transparency and access concerning influenza virus genetic sequence data was also recognized by the 64th World Health Assembly in 2011.¹⁰⁰ Keiji Fukuda, in his capacity as Assistant Director-General for Health Security at the WHO at the time, described GISAID as a “critically important and technically advanced new platform” that “provides an important option for sharing genetic sequence and epidemiological data” (Fukuda, 2011). The hosting of the database and platform by the German government is also seen to provide an important component of the trust GISAID now enjoys in many countries (ibid.). A final, but critical, component to maintaining this trust is GISAID's declaration that – since its formation – neither the initiative, its management, nor its board members have received research support, investment interests, or performed any contract work for industry or other commercial entities.¹⁰¹

Conclusion

GISAID may have had an unlikely birth as a new global health initiative – with an unusually strong role played by an energetic, influential, and dedicated philanthropist without a prior background in global health. As the initiative marks its 10th anniversary, however, it is evident that GISAID is now making significant contributions to global health. Five such contributions stand out: (1) collating the most complete repository of high-quality influenza virus data; (2) facilitating the rapid sharing of potentially pandemic virus information during recent outbreaks; (3) informing the WHO's biannual seasonal flu vaccine strain selection process; (4) developing informal mechanisms for conflict resolution; and (5) building greater trust with low-income and middle-income countries key to pandemic preparedness. Indeed, an array of interviewees pointed out that there is now widely perceived merit in GISAID's formula for balancing the need for control *and* openness, as well as the way it seeks to reconcile the competing imperatives of science, public health, and business.¹⁰² A number of interviewees within and outside of GISAID variously felt that it is functioning well as a mechanism and proving its value,¹⁰³ that

⁹²Interview with Yuelong Shu.

⁹³PIP Framework Advisory Group, 'Draft optimal characteristics of an influenza genetic sequence data sharing system under the pip framework', (2015) Available at: http://www.who.int/influenza/pip/advisory_group/draft_twg_doc.pdf [Accessed 3 December 2015]; Interview with Ron Fouchier.

⁹⁴Interview with a European expert in research and biotechnologies issues, formerly working in the European Commission. 19 February 2015; Interview with Hanns-Christoph Eiden, President of the Federal Office for Agriculture and Food, Bonn. 20 February 2015.

⁹⁵Search of EpiFlu database performed on 7 January 2016.

⁹⁶Search of EpiFlu database performed on 7 January 2016.

⁹⁷Email correspondence with GISAID Initiative (Freunde von GISAID e.V.). 8 January 2016.

⁹⁸GISAID Training Workshops, https://isirv.org/site/images/stories/avg_documents/Cape_Town/AVG-GISAID%20Workshop%20Report_OptionsVIII.pdf; <https://isirv.org/site/images/conferences/NGS/NGS%20Workshop%20Programme.pdf>; <https://www.isirv.org/site/images/Report%20on%20GISAID%20Workshop%20HK.pdf>; <https://www.ttsh.com.sg/page.aspx?id=7014>; <http://www.influenza.spb.ru/en/conferences/gisaid-2014-en>. [All accessed 23 March 2016].

⁹⁹CIDRAP News, 'Pandemic reveals strengths of new flu database' (2009) Available at: <http://www.cidrap.umn.edu/news-perspective/2009/06/pandemic-reveals-strengths-new-flu-database> [Accessed 26 November 2015].

¹⁰⁰Sixty-Fourth World Health Assembly. Geneva, 16–24 May 2011. Resolutions and Decisions Annexes. Available at: http://apps.who.int/gb/ebwha/pdf_files/WHA64-REC1/A64_REC1-en.pdf. [Accessed 14 January 2016].

¹⁰¹PIP Framework Advisory Group, 'Draft optimal characteristics of an influenza genetic sequence data sharing system under the pip framework', (2015) Available at: http://www.who.int/influenza/pip/advisory_group/draft_twg_doc.pdf. [Accessed 3 December 2015].

¹⁰²Interview with Philip Dormitzer.

¹⁰³Interview with David Heymann, Centre for Global Health Security, Chatham House, 9 June 2015.

there is a clear need for it,¹⁰⁴ that it is (successfully) contributing to changing habits around virus sharing,¹⁰⁵ and that it has now effectively become the “go-to” source for influenza information, especially when new outbreaks happen.¹⁰⁶

That said, even today it remains hard to know just how much genetic sequence data are still *not* being shared. Without knowing how many influenza viruses are being sequenced internationally (and when they are being sequenced), it is simply impossible to tell.¹⁰⁷ Some countries evidently prefer to share influenza data through GISAID rather than through other platforms that do not provide contributors with the same levels of protection yet some interviewees also expressed suspicion that – especially in the area of animal health – there still is much information that is not being shared, and that some countries are still only sharing a small proportion of information.¹⁰⁸

Although GISAID has developed a distinct ethos for sharing, moreover, there are also natural limits regarding its ability to ensure that the terms of its access agreement are adhered to. Limiting access – and even outright exclusion – of those who violate the terms of the access agreement remains a credible sanction, and one that has been used in the past. According to GISAID, the percentage of all active users whose access credentials to the GISAID platform have been revoked at the time of writing is around 0.16%.¹⁰⁹ At present, however, GISAID is only able to trace who is accessing information, not whether people are passing this information on to others.¹¹⁰ GISAID maintains that if such data subsequently surfaces, they do have means to prove someone has illegally obtained the data – meaning that data contributors who suspect violations could seek to pursue this through legal channels.¹¹¹ Nor, of course, can there be ultimate guarantees that people will adhere to these rules when confronting all the pressures of a pandemic situation in future, although the initiative has now developed a successful track record of navigating such situations.¹¹²

Looking forward, there are at least three challenges that GISAID will need to navigate over the medium to long term. One such longer-term challenge mentioned by several respondents revolves around the future leadership of the initiative. On the one hand, the material sustainability of the initiative is ensured for the foreseeable future through vital support provided by the federal government of Germany.¹¹³ On the

other hand, it is evident that the personal investment of significant amounts of time, energy, expertise, skill, and commitment by its philanthropic champion has been central to the success of GISAID.¹¹⁴ The longer term question of who will continue to provide wider leadership and the championing of GISAID will have to be addressed. Towards this end, the German government recently called for the institutionalization of GISAID to ensure its longevity.¹¹⁵

A second challenging area for GISAID are the ongoing negotiations around the Pandemic Influenza Preparedness (PIP) Framework. The PIP framework creates responsibilities for the sharing of biological samples of influenza viruses with pandemic potential as well as providing a mechanism – partially funded by industry contributions – for the provisions of benefits (like medicines and vaccines) to affected countries. The question of whether genetic sequences data (as opposed to physical specimens) should also be governed by the framework proved too sensitive to be resolved during the initial negotiations for the PIP framework (Gostin et al., 2014). At the time of writing, GISAID is thus having to navigate a complex and sensitive set of diplomatic negotiations around the future role of genetic sequence data in the framework, with potentially considerable ramifications for the future of the initiative.

Probably, the biggest question to arise from GISAID’s success, however, is whether its sharing mechanism can be extended to also cover other viral diseases and stand for a wider paradigm shift in improving international data sharing. As we have already seen several times over the past decade, influenza is not the only potentially lethal infectious disease that the world confronts, and it remains to be seen whether GISAID has both the aspiration and capacity to expand into other lethal viral diseases – such as Ebola, MERS, West Nile Virus, and other zoonotic diseases.¹¹⁶ There are certainly signs that this data sharing problem also manifested itself during the more recent Ebola outbreak in West Africa, (Yozwiak et al., 2015) as well as the spread of Zika virus.¹¹⁷ Although GISAID was approached to extend its scope to include Ebola viruses, GISAID’s leadership felt at the time that this was beyond its capacity. Yet Ebola is clearly another disease that shows more generally that the spread of viruses and their husbandry systems do not align naturally with political boundaries.¹¹⁸ Here too, as in a number of other lethal infectious diseases,

¹⁰⁴Interview with Nancy Cox.

¹⁰⁵Interview with Alan Hay.

¹⁰⁶Interview with Philip Dormitzer.

¹⁰⁷Interview with Ron Fouchier.

¹⁰⁸Interview with Gwenaëlle Dauphin.

¹⁰⁹Email correspondence with GISAID Initiative (Freunde von GISAID e.V.). 18 January 2016.

¹¹⁰Interview with Masato Tashiro.

¹¹¹Email correspondence with GISAID Initiative (Freunde von GISAID e.V.). 18 January 2016.

¹¹²Interview with John McCauley.

¹¹³Interview with an expert in computational biology. 23 February 2015.

¹¹⁴Interview with Ron Fouchier; Interview with an expert in computational biology. 23 February 2015; Interview with John McCauley.

¹¹⁵Minutes from the Annual General Meeting between Germany and GISAID. 28 April 2015.

¹¹⁶Interview with an expert in computational biology. 23 February 2015; Interview with Ron Fouchier.

¹¹⁷‘Brazil Considers Reforming Biosecurity Law Amid Criticism’. Associated Press. 6 February 2016. Available at: <http://www.ndtv.com/world-news/brazil-considers-reforming-biosecurity-law-amid-criticism-1274312>. [Accessed 10 February 2016].

¹¹⁸Interview with Ian Brown.

protecting populations will require creative and sustained efforts to carefully reconcile data, disease, and diplomacy.

Conflict of Interest

The authors have declared no conflict of interest.

Author Statement

S. Elbe and G. Buckland-Merrett jointly conceived and designed the study, developed the study protocol, collected the data, analysed the data, and prepared and approved the manuscript.

References

- Bogner, P. et al. 2006. "A global initiative on sharing avian flu data". Available at: <http://www.nature.com/nature/journal/v442/n7106/full/442981a.html>. [Accessed 21 December 2015].
- Brown, D., 2006. 'Bird Flu Fears Ignite Debate on Scientists' Sharing of Data'. The Washington Post. 25 May 2006. Available at: <https://www.washingtonpost.com/archive/politics/2006/05/25/bird-flu-fears-ignite-debate-on-scientists-sharing-of-data/ca5ce4e-c68e-44bd-bc07-6e423a58475a/>. [Accessed 11 January 2016].
- Butler, D., Cyranoski, D., 2013. 'Flu papers spark row over credit for data'. Nature, 497, 14–15.
- Fukuda, K., 2011. Comments at 'Breaking through the InfluenzaWalls – A Symposium Hosted By The Federal Office For Agriculture And Food'. Bonn, Germany. 11 October 2011. Available at: <https://www.youtube.com/watch?v=M5ekNjxuArk>. [Accessed 11 January 2016].
- Gostin, L.O., Phelan, A., Stoto, M.A., Kraemer, J.D., Reddy, S., 2014. 'Virus sharing, genetic sequencing, and global health security'. Science 345(6202): 1295–1296.
- Greenemeier, L., 2009. 'Open-access flu research web site is relaunched amid controversy', Scientific American (2009). Available at: <https://www.scientificamerican.com/article/gisaid-sib-flu-database>. [Accessed 25 November 2015].
- Longo, D., Drazen, J., 2016. 'Data Sharing'. New England Journal of Medicine 374(3), 276–277.
- Neher, R., Bedford, T., 2015. 'nextflu: real-time tracking of seasonal influenza virus evolution in humans', (2015) Bioinformatics. Available at: <http://bioinformatics.oxfordjournals.org/content/early/2015/07/03/bioinformatics.btv381>. full.pdf. [Accessed 4 December 2015].
- Pearson, H., 2006. 'Plan to pool bird-flu data takes off'. Nature 442. 31.
- Quick M., 2006. "Non-WHO global initiative on sharing avian influenza data". The Lancet Infectious Diseases 6(10): 621. October.
- Roos, R., 2006. 'Scientists launch effort to share avian flu data'. CIDRAP. Available at: <http://www.cidrap.umn.edu/news-perspective/2006/08/scientistslaunch-effort-share-avian-flu-data>. [Accessed 26 January 2016].
- Schnirring, L., 2009. 'German government to host flu database'. CIDRAP. 20 October 2009. Available at: <http://www.cidrap.umn.edu/news-perspective/2009/10/german-government-host-flu-database> [Accessed 23 March 2016].
- Sedyaningsih, E.R., Isfandari, S., Soendoro T., Supari, S.F., 2008. 'Towards Mutual Trust, Transparency and Equity in Virus Sharing Mechanism: The Avian Influenza Case of Indonesia'. Annals of the Academy of Medicine (Singapore). June. 37(6): 482–488.
- Supari, S., 2008. *It's time for the world to change; in the spirit of dignity, equity, and transparency*. Sulaksana Watinsa Indonesia (SWI), Jakarta.
- World Health Organization. 2016. 'FAQs H5N1 Influenza' Geneva. Available at: http://www.who.int/influenza/human_animal_interface/avian_influenza/h5n1_research/faqs/en/ [Accessed 07 January 2016].
- Yozwiak, N.L., Schaffner, S.F., Sabeti, P.C., 2015. 'Data sharing: Make outbreak research open access'. Nature. 25 February 2015. Available: <http://www.nature.com/news/data-sharing-make-outbreak-research-open-access-1.16966>. [Accessed 11 January 2016].
- Zamiska, N., 2006a. 'Scientist Rebels Against WHO Over Bird Flu'. The Wall Street Journal. 13 March 2006. Available at: <http://www.wsj.com/articles/SB114221407089396217>. [Accessed 11 January 2016].
- Zamiska, N., 2006b. 'A Nonscientist Pushes Sharing Bird-Flu Data', The Wall Street Journal. 30 August 2006 Available at: <http://www.wsj.com/articles/SB115696782401349781> [Accessed 23 March 2016].

Submission Date: 10/20/2017

Name: Toby Bloom

Name of Organization: New York Genome Center

Type of Organization:

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

It is absolutely critical to progress in genomic research that we enable the easiest, broadest access to genomic data, subject to adequate protection of subjects. Summary data, inherently less risky than access to individual-level data, should be made as widely accessible as possible.

The click-through agreement, as defined in the current proposed policy, provides a compromise between full public access to summary data and the current controlled-access model. The proposed click-through solution raises questions as to whether scientists will be able to access data easily enough, and whether they will have access to a broad enough slice of all summary data (the rest being classified as sensitive). On the privacy side, questions have been raised about whether such a click-through, requiring data users to agree not to attempt to find data that violates participants' privacy (wording not exactly clear yet), provides adequate protections to those genomic subjects.

Daniel McArthur [<https://macarthurlab.org/2017/10/10/response-to-proposal-to-update-data-management-of-genomic-summary-results-under-the-nih-genomic-data-sharing-policy/>] has objected to the NIH proposal, saying that click-throughs on each access are overly burdensome to researchers, especially for automated access, and provide no significant protection, since everyone is accustomed to clicking-through terms of access all the time. He proposes that access to summary data should be publicly available. Yaniv Erlich [<https://medium.com/@erlichya/why-clickthrough-agreement-is-the-right-approach-for-aggregate-genetic-data-response-to-the-a8523408b390>] has countered that the threat to privacy is real, and increasing, as the technology for these differential privacy attacks has improved, and thus protections are still warranted. He further argues that there are technologies for making click-through less burdensome for API access.

It is true that most users are fully conditioned to simply click-through conditions without reading them, and thus the concern that this mechanism won't provide any additional data protection is valid. It is also true that requiring a click-through every time a user wants to access data is at least annoying to researchers, and may pose additional burdens for automated access. But eliminating the click-through entirely, as has been proposed,

makes the data publicly available with no explicit conditions on appropriate use of the data, and no consequences for misuse. We would argue that some enforcement of appropriate use is warranted.

The Global Alliance has proposed a somewhat different mechanism - “registered users”. They propose a scenario in which a researcher registers ONCE, not each time they use the data. They provide an email address (and possibly a phone number as well), for a valid research institution and agree to the terms of use. Then either NIH simply validates the address by sending an email (and text) that require coded responses, or, if they want to restrict usage to researchers only, may require an institutional signing official to validate the researcher, much the way they currently validate NIH eCommons accounts. The primary objection to institutional approval is that it might make it difficult for citizen scientists to work on the data. That is a decision NIH must make. In either case, the user would then be provided an ID or a token, with which they could access the data either via API or web access at later times. Yaniv Erlich has provided details for one such API solution. This reduces the burden on scientists to a one-time registration, while still providing more protection than click-through and makes the data available more easily for research. It also provides a mechanism for implementing consequences for violation of the agreement. We think the simpler email validation (or additionally phone #, two-factor authentication), is a reasonable compromise, and removes the administrative burden for researchers. It doesn’t remove all risk: It does not eliminate the possibility that malicious users could game this system by providing a temporary email address or phone number, but malicious users can breach systems containing individual data also.

One caveat: if there is any incident in which a participant experiences harm from this release of summary data, and that incident gets any level of publicity, the impact on future research projects and the precision medicine initiative will likely be severe – people will become much more reluctant to participate, and members of isolated or under-represented communities, who are more distrustful of our institutions in general, will likely be disproportionately affected. We need to be sure that the measures we take here are sufficient to minimize that risk.

2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity

Maintaining summary information for sensitive data under controlled-access is certainly better than not providing such summary data at all. But I would propose that there are techniques for making some subsets of that data more openly available - maintaining only some aggregations in controlled access, and making aggregates that do not disclose sensitive data available. Datasets do not necessarily partition into those that are sensitive and those that are not. For example, if an isolated population is part of a large study, summaries by population may be sensitive, but summaries by disease might not be (unless the disease incidence correlates with that population - this requires a statistical analysis of the correlations.) Avoiding differential privacy attacks, such as Homer [<http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000167>] and

Bustamante[<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4667107/>], requires more extensive analysis to determine, for example, how many successive queries can be permitted without risk of leaking personal information, or which summaries require restricted access. But dividing datasets into sensitive and not sensitive seems to be too coarse-grained a criterion. There is more work to be done here. We can start with leaving those studies under controlled-access, but NIH should be investing in making those data more widely available, to the greatest extent possible.

3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive

This is perhaps the area that raises the most concern in the proposal. It is not clear that those submitting institutional certification will have adequate understanding of the sensitivity or risk in the data they are providing. Many of the risks inherent in summary data are related to statistical risks dependent on the size and distribution of the data in various aggregation/stratification scenarios. That likely means that some institutions will err on the side of safety and declare data to be sensitive when it is in fact low risk, while other organizations will assume summary data to be safe when it carries higher risks. In one case, researchers will not get the access they need to large amounts of data. In the other, we risk causing a real or perceived harm, and any wide publicity about such a privacy concern could have broad implications for the willingness of other participants to agree to contribute their genomic data to science.

Determination of the sensitivity of data is a non-trivial problem. The policy might be better served by asking, not just whether data is sensitive, but rather asking a set of detailed questions about the datasets as part of the institutional certification, that could then allow a more focused data access committee to resolve this question. Creating that list of questions is beyond this set of comments, but I think could be done by an expert privacy committee.

4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies

Submission Date: 10/20/2017

Name: Tejia Zhang

Name of Organization: University of Utah

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Please see attached document.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Please see attached document.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

Please see attached document.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Please see attached document.

Additional Comment (attachment):

Request for Comments: Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy

Response to question 1

NIH's proposal¹ for inclusion of genomic summary data in a new rapid access tier will yield significant benefits, particularly for those engaging in research. Similar to abstract or research summary provided at the beginning of most scientific publications, the rapid access of summary statistics will give researchers a quick overview of the entire study, and aid them in determining if more detailed information need to be further accessed. This rapid access of genomic summary data will allow researchers to browse through larger numbers of projects in a shorter amount of time, and improve research efficiency.

The availability of genomic summary data is not a new concept, as statistical information such as standard errors and p-values are often included in publications, thereby already viewable by all researchers with access to the journals in question. Genomic summary results have also been made public by the research community in resources such as the Genome Aggregation Database² and Type 2 Diabetes Knowledge Portal.³

In considering the benefits associated with the proposed policy change, it is also crucial to address potential risks, which include breach of patient privacy and unintentional release of personal health data. As is mentioned in NIH's proposal,¹ Homer *et al.* demonstrated the inability of summary statistics to fully mask individual-level data, as it is possible to predict an individual's participation in a research study *via* a combination of summary statistics and full genome sequence.^{1,4} This is no doubt detrimental both in terms of breach of privacy, as well as unwanted release of sensitive health information that could be used negatively against a patient, family members, and/or other members of the patient's community.¹ While an actual instance of this has not occurred,^{1,5} and the requirement for simultaneous access of summary statistics and large amount of SNP data suggests that the current risk is low,^{1,4} this issue highlights the importance of patient privacy and data safety, and the indispensability of security measures such as de-identification and separate storage for portions of data that could be merged to recover identity.

In addressing safety concerns, a stricter classification of genomic summary data into more difficult-to-access tiers may not be the optimal solution, as this may unnecessarily slow scientific discoveries, which could adversely affect the patient populations these discoveries are meant to serve. Classification of summary data under controlled access could also prove counterproductive toward data security, as researchers who require only summary statistics will need to apply for access to a more restricted tier that may also contain individualized genotypic and phenotypic information;¹ creation of a rapid access tier for genomic summary data will alleviate this issue by reducing unnecessary requests for controlled access.

As updates are made to accommodate the ever-changing landscape of genomic research, more efforts also need to be invested in making available trainings and information on data security, informed consent, and the ethical issues associated with the rise of genomic technologies, especially for researchers newly entering this field. As practiced in a number of academic institutions, researchers awarded genomics-related NIH training grants are required or recommended by the granting department to partake in genomic medicine/medical ethics courses, which could prove beneficial in preparing these researchers for future projects in genomics.

Response to question 2

As is stated in response to question 1, NIH's proposal for a rapid access tier for genomic summary data will greatly facilitate data access and improve research efficiency. On the other hand, as is stated in the proposal,¹ the findings of Homer *et al.* support the capability of reconstructing an individual's participation status in a study based on summary statistics and that individual's whole genome sequence.⁴ While current risk appears low given the accuracy of prediction depends on availability of large amount of SNP data,⁴ this could become more concerning as sequencing cost decreases and whole-genome data become more readily available. This issue is particularly worthy of consideration for studies involving populations from isolated geographic locations, small patient populations with rare genetic diseases, or information that could be potentially stigmatizing if released.¹

It is also possible that the creation of an access tier specifically for summary statistics could welcome deposition of summary data into that tier even in instances where sensitivity of the study supports inclusion of summary data under controlled access as more appropriate. To reduce this risk, it is crucial that institutions depositing data be made aware of their choice to continue submission of summary data into the controlled tier if they deem that the more reasonable option; this also speaks to the importance of web portal design to maximize clarity and facilitate data deposition into the intended tiers.

Response to question 3

The question of who holds authority in the designation of data as sensitive is a complex one, and needs to be assessed from different angles. On one hand it could be argued that researchers are the most familiar with their own data, and would be the most qualified in determining what information could be potentially detrimental to patient communities if released rapidly. The diversity of projects under human genomics also suggests that a single set of regulations governing data designation may not be appropriate in all contexts, and could lead to unnecessary slowing of data release.

On the other hand, the current proposal^{1,6} does not encompass significant oversight, and leaves open potential for unintentional or intentional policy breach. While one possible solution is creation of a committee either at the level of individual institutions or NIH specifically for the evaluation of each project's level of sensitivity prior to repository submission, this may not be optimal in practice. Significant variability exists across genomic research, both in the communities participating and technologies used, while recent improvements in sequencing and increased knowledge of genomics' capabilities may push already-existing projects toward this direction, and lead to the evolution of older projects to incorporate genomic technologies to varying degrees. While all these factors confound the questions of who holds responsibility in the assignment of sensitive data and what criteria should be used in these evaluations, their complexity suggests that a one-size-fits-all approach may not be the ideal solution. Additionally, the sheer number of projects under genomics renders a committee-based approach time-consuming and costly to execute.

For the time being, an alternative that balances the need for patient protection with the freedom of data exchange could be to increase overall awareness of the many still unaddressed ethical issues associated with this evolving field, by including discussion of genomics as a part of seminars and courses at the institutional level, and making it possible for investigators to efficiently communicate with and access help from contacts at the NIH. Given the already well-

established presence of review systems such as IRB⁷ and IACUC,⁸ it may be possible for members of these groups to take a more active role in providing recommendations in the realm of data sensitivity. Similar to the click-through agreement associated with rapid access, it could also be beneficial to include more data sensitivity-related click-through information (for example, a reiteration of NIH's current tier system, and a mentioning of the types of studies mostly likely to include sensitive data, such as studies of populations from isolated geographic regions or rare genetic diseases¹) at the level of data submission to repositories, as a reminder of the importance of careful designation of research data.

Response to question 4

While not directly related to the rapid access proposal detailed in this current update, I would like to bring up three additional comments related to GDS as follows:

1. Sharing of non-human genomic data

Under current GDS policy, a six-month release timeline is given following initial data cleanup for level 2-4 human genomic data, while non-human genomic data release is expected no later than date of publication.^{6,9} While it could be argued that direct relevance of human genomic data to health and therapeutic development supports this earlier release of human data, it is also worth noting the importance of animal models and sequence information for key model organisms (as an example, assembly of the zebrafish genome¹⁰⁻¹¹ paved the way for successful modeling of a number of human diseases in this organism,¹² and high-throughput chemical screens in zebrafish larvae have yielded lead compounds that are currently undergoing optimization for treatment of conditions including long QT syndrome and fibrodysplasia ossificans progressiva¹²). Given the importance of model organisms in elucidation of disease pathways and therapeutic development, the requirement for non-human genomic data release only at the point of publication could cause unwanted delays in research that often lies upstream of clinical trials and patient studies; this could be particularly problematic for newer model organisms with significant potential in human health, but still require more updates in their genome assembly.¹³⁻¹⁴ As an alternative to the current policy, the six-month human data release timeline could also be adopted for non-human data, which could result in more clear, streamlined data release instructions and reduce delays in model organism research.

2. Comparing six-month hold against an embargo-based approach

While the six-month data holding period could be viewed as a reasonable middle ground between immediate data release accompanied by loss of proprietary lead-time, and longer hold times that significantly delay research progress,¹⁵ it could still be argued that immediate availability of genomic data without hold time will significantly improve efficiency for researchers whose work could benefit from the newly released data.¹⁵ In this sense, an embargo approach may better achieve the compromise between immediate data availability and lead-time for generators of the original data.¹⁵ While enforcement of the embargo could be challenging due to the large number of researchers accessing data,¹⁵ it may be achievable through measures such as clear instructions (i.e. reiteration of embargo policy and asking for affirmation from researchers for honest upholding of embargo) at the stages of data access and publication.

3. Fine-tuning data release policies for additional -omics-based data

Just as improvements in sequencing have significantly increased the bandwidth of genomic research and prompted the updating of genomic data-sharing policies, perhaps similar protocols should be considered for other -omics-based approaches. In particular, advancements in mass spectrometry-based technologies have significantly reduced analysis time while boosting resolution and sensitivity,¹⁶ allowing quantification of thousands of metabolites from a single sample within a relatively short amount of time,¹⁶ and implementation of high-throughput, comparative metabolomic studies across different patient populations.¹⁷ While contributions to publically available databases such as LIPID MAPS¹⁸ and METLIN¹⁹ have greatly facilitated metabolite identification and data analysis, there does not appear to exist consistent recommendations for deposition of large-scale metabolomic data into designated repositories. As more metabolomics and other -omics-based data are generated, perhaps it is time to begin consideration of more concrete guidelines for the timely uploading of these data to repositories to promote more efficient data sharing across the scientific community. While still partly a goal for the future, one could envision the benefits of cross-referencing genomic data with other omics-based information to more precisely map genotype-phenotype correlations;²⁰ accomplishing this goal will require the availability of comprehensive -omics data repositories and robust network infrastructure supporting communication across databases.²⁰

References

1. Request for comments: proposal to update data management of genomic summary results under the NIH genomic data sharing policy [website online]. Bethesda, MD: National Institute of Health; 2017. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-110.html>. Accessed October 20, 2017.
2. Genome Aggregation Database (gnomAD) [database online]. Cambridge, MA: Broad Institute; 2017. <http://gnomad.broadinstitute.org>. Accessed October 19, 2017.
3. Type 2 Diabetes Knowledge Portal (T2DKP) [database online]. Bethesda, MD: NIH Accelerating Medicines Partnership; 2017. <http://www.type2diabetesgenetics.org>. Accessed October 20, 2017.
4. Homer N et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet.* 2008;4(8):e1000167. doi:10.1371/journal.pgen.1000167.
5. Wendler DS, Rid A. Genetic research on biospecimens poses minimal risk. *Trends Genet.* 31(1):11-15. doi:10.1016/j.tig.2014.10.003.
6. NIH genomic data sharing policy [website online]. Bethesda, MD: National Institute of Health; 2014. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html>. Accessed October 20, 2017.
7. Research Involving Human Subjects [website online]. Bethesda, MD: National Institute of Health; 2016. <https://humansubjects.nih.gov>. Accessed October 20, 2017.
8. The Institutional Animal Care and Use Committee [website online]. Bethesda, MD: National Institute of Health; 2011. <https://grants.nih.gov/grants/olaw/tutorial/iacuc.htm>. Accessed October 20, 2017.
9. Supplemental Information to the GDS Policy [website online]. Bethesda, MD: National Institute of Health; 2014. https://osp.od.nih.gov/wp-content/uploads/Supplemental_Info_GDS_Policy.pdf. Accessed October 20, 2017.

10. Howe K et al. The zebrafish reference genome sequence and its relationship to the human genome. *Nature*. 2013;496(7446):498-503. doi:10.1038/nature12111.
11. Zebrafish Genome Project [database online]. Hinxton, Cambridgeshire, UK: Wellcome Trust Sanger Institute; 2017. <http://www.sanger.ac.uk/science/data/zebrafish-genome-project>. Accessed October 20, 2017.
12. Rennekamp AJ, Peterson RT. 15 years of zebrafish chemical screening. *Curr Opin Chem Biol*. 2015;24:58-70. doi: 10.1016/j.cbpa.2014.10.025.
13. Russell JJ et al. Non-model model organisms. *BMC Biol*. 2017;15(1):55. doi:10.1186/s12915-017-0391-5.
14. Goldstein B, King N. The future of cell biology: emerging model organisms. *Trends Cell Biol*. 2016;26(11):818-824. doi:10.1016/j.tcb.2016.08.005.
15. Contreras, JL. NIH's genomic data sharing policy: timing and tradeoffs. *Trends Genet*. 2015;31(2):55-57. doi:10.1016/j.tig.2014.12.006.
16. Johnson CH, Ivanisevic J, Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol*. 2016;17(7):451-459. doi:10.1038/nrm.2016.25.
17. Gao X et al. Large-scale metabolomic analysis reveals potential biomarkers for early stage coronary atherosclerosis. *Sci Rep*. 2017;7(1):11817. doi:10.1038/s41598-017-12254-1.
18. Fahy E, Sud M, Cotter D, Subramaniam S. LIPID MAPS online tools for lipid research. *Nucleic Acids Res*. 2007;35(suppl 2):W606-W612. doi:10.1093/nar/gkm324.
19. Smith CA et al. METLIN: a metabolite mass spectral database. *Ther Drug Monit*. 2005;27(6):747-751. <https://insights.ovid.com/pubmed?pmid=16404815>. Accessed October 20, 2017.
20. Wanichthanarak K, Fahrman JF, Grapov D. Genomic, proteomic, and metabolomic data integration strategies. *Biomarker Insights*. 2015;10(suppl 4):1-6. DOI: 10.4137/BMI.S29511.

Submission Date: 10/20/2017

Name: Goncalo Abecasis

Name of Organization: University of Michigan

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

The planned update is a large and substantive improvement to NIH's policy for sharing of genomic summary statistics. In our view, the policy has several important and desirable features including a broad definition of per-marker summary statistics (which includes allele frequencies, but also effect sizes, standard errors, and p-values) and the recognition that the controlled access process should be different depending on the potential risk and harms to research participants (which are much higher when individual level data and individual level data with HIPAA protected identifiers is shared than when summary statistics are shared). In this way, the update will accelerate research, reduce the need for creating and re-analyzing duplicate copies of individual level data, and encourage new research methods and tools that use summary statistics to facilitate new insights. We applaud the NIH's decision to increase use of the valuable data and to accelerate biomedical research, which is consistent with the intentions of the donors of the data and will aid scientific progress and, ultimately, public health.

The proposed click-through mechanism has several appealing features, particularly in minimizing the likelihood that summary statistics will be used in proof-of-principle attacks that demonstrate technical strategies for re-identifying research participants. However, it also has the potential to reduce re-use of the data because (for example) it is not clear whether the possibility allows users to combine and remix the data (for example, by further aggregating statistics across studies) and the redistribute the results of those analyses. It would be important to clarify that third parties should be able to carry out these downstream analyses and redistribute results, perhaps with the proviso that the same click-through agreement should apply to downstream uses. Many in the genomic community would have preferred an even more open policy, that does not require a click through agreement. We strongly encourage NIH to consider such a policy especially for analyses that have the lowest levels of risk, that is those that (a) use only genomic information but not attached phenotypes (as when allele frequency databases are created) and (b) those where the number of independent markers is smaller than the number of subjects being analyses (that situation protects against attacks aimed at re-identification as well as against attacks aimed at reconstructing individual phenotypes after re-identification).

2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity

The decision to classify information as sensitive should be seconded by an independent and non-affiliated body of qualified assessors. For highly sensitive phenotypes (for example, those relating to consumption of illegal drugs or behaviors that are subject to extreme social stigma) and analysis of especially sensitive populations (particularly disadvantaged minority groups that are most under-represented in research), we agree it is prudent to retain controlled access for now. Those are the situations where harms from potential re-identification seem the most likely and also where it is essential for NIH to focus on building trust among potential research participants. However, we strongly encourage NIH to consider -- to the extent possible -- clear definitions and a limited set of examples of when these exceptions might be claimed. That will minimize inconsistencies, clarify participant expectations, and avoid situations where the exception might be claimed inappropriately.

3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive

As noted above, we strongly encourage NIH to consider -- to the extent possible -- clear definitions and a limited set of examples of when these exceptions might be claimed. It would also be ideal to maintain a list of examples of traits and populations where exceptions are not considered appropriate. That will minimize inconsistencies, clarify participant expectations, and avoid situations where the exception might be claimed inappropriately.

4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies

The portion of the policy that relates to changes in informed consent is somewhat concerning. Ongoing research on informed consent (see for example Roessler et al, 2015, PMC4382351) strongly suggests that simplifying the informed consent process is extremely important to both increase participant understanding of research risks and benefits and to increase their overall trust in the research process (which sometimes actually entails reduced rates of participation). We are concerned that requiring specific language in consent forms requiring discussion of sharing of aggregate data and statistics and other mandates for specific individual items that must be discussed and explained in consent forms risks making these unwieldy, less likely to be read or understood by participants, and ultimately does not fulfill the ultimate goal of making participants more informed about the research they are donating samples, time, and data to. In our view, it should not generally be necessary to discuss sharing of marker, gene or region level

summary statistics (that aggregate information across thousands of individuals) when participants have already consented to sharing of individual level data since the process entails no extra risk. In exceptional studies that promise participants their data will not be shared, it may be appropriate to explain that aggregate or summary statistics that synthesize information across large groups will be shared. Requiring consent for uses of aggregate data also blurs an important distinction between individual level data (which is associated with and sometimes clearly identifies specific individuals) and summary statistics which have long been published in research papers, websites, conference proceedings, and many other contexts by the scientific community (albeit at a smaller scale). These statistics effectively synthesize knowledge, which should typically be broadly shared and reused.

We also believe that the policy could be improved by allowing some types of summary statistics, particularly those derived without the use of disease status information (such as allele frequencies derived from very large cohort or consortium studies) and those derived from extremely large studies should be shared even more broadly, without requiring click-through agreements. For example, once data is aggregated in very large numbers of individuals (operationally, this typically will include datasets where the number of subjects exceeds the number of independent markers being analyzed, $N > M$), the risk of re-identification is greatly minimized as is the very risk of phenotype reconstruction after re-identification. We encourage NIH to solicit community input on an appropriate definition of very large (perhaps, $N > 500,000$) and to make a simplified access tier at least for those aggregate / summary results. For example, such statistics (derived from very large datasets) could be made available systematically without click through requirements. Such statistics are both highly valuable and likely to become much more common in the future as data aggregation becomes simpler and large population studies become more common.

(beyond current language describing the desire for sharing individual level genomic data and phenotypes)

Submission Date: 10/20/2017

Name: Nancy Cox

Name of Organization: Vanderbilt University Medical Center

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

The risks are so modest as to have little real meaning when weighed against the long-term value to patients of more scientists working on more datasets.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

I think there are a more limited number of studies that should qualify as sensitive than has been described. Universities are so lawyer driven, they will be way too conservative about what should be sensitive. They protect the University not patients. There are genuinely sensitive datasets, but they are really rare, and the suggested approach is too conservative.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

All published papers should be required to submit all results. But don't bother with making NCBI calculating results on datasets. No one will use that. The results from published papers would be used. The NCBI results, probably not.

Submission Date: 10/20/2017

Name: T Degui Zhi

Name of Organization: University of Texas health science center at houston

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Benefits outweigh risks.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

This will slow down the pace of discovery. It will create hurdles for software tools that can effectively leverage the big data sets. It's not a good use of publicly funded resources.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 10/20/2017

Name: Dr. Manuel A Rivas

Name of Organization: Stanford University

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Additional Comment (attachment):

October 20, 2017

Response to “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy”

We would like to thank NIH for moving forward with a “Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy”.

We want to make especially clear to NIH and the public our position that we truly accept that sharing of genomic summary results has become crucial for scientific and clinical discovery. In both the Rivas and Bustamante labs these data have accelerated discoveries that will transcend our understanding of human health and disease.

Absence of informed risk can lead to asymmetric information between an individual and a negotiating party, e.g. an insurance company, which can create moral hazards by increasing exposures to risk, e.g. bankruptcy. These are not academic thought scenarios, but rather real life scenarios which we have either dealt with or have family and friends that have been exposed to such imbalance, e.g. predatory lending.

Hence, we would like to take this opportunity to comment on the risks we foresee with the **willingness to distribute** summary statistic data, but also the **hesitation to distribute** summary statistic data from particular individuals and communities as we enter the 21st century of human genomic medicine research. Finally, we would like to enumerate some solutions and how they align with the proposed access mechanisms by NIH.

Risks with willingness to distribute summary statistic data

1. Increased risk for utilizing data to introduce harm to vulnerable communities and populations

In a world where access to genome sequencing technology is universal the answer is clear that as more information is gained from large-scale genomic summary level data combined across hundreds of medical traits the barrier to reidentification is lowered. The most susceptible individuals to what may be a rare event (reidentification) will likely be the most vulnerable to negative consequences, i.e. those that either have limited to no health coverage due to either pre-existing conditions or perceived pre-existing conditions. As we improve our understanding of human disease, the more we will learn about the interplay between genetics and the environment and the effect it has on disease onset and progression. For many outcomes, research suggests that that the environment will exacerbate genetic risk. Thus, less privileged and marginalized communities, which are disproportionately affected by environmental factors, may be the first in line where this information is used to negatively impact health coverage decisions.

2. Increased risk to principal investigators and institutions who may be most vulnerable in an attack

In the event of an unlikely attack, undue blame can be placed on the principal investigators (PIs). Without explicit mandates against abuse of summary statistics, if a malicious user abuses the available data the public may see that as an irresponsible handling of the data by the PIs. This negative optics not only can compromise the particular PI and institution but may discourage future participants and negatively affect the entire field. Additionally, without proper guidelines from NIH to mitigate risks for summary statistic dissemination, there is increased risk for PIs who are willing to share summary statistic data, especially if the PI comes from less privileged and marginalized communities. Furthermore, at a less privileged university, without NIH guidelines, the university staff might not be adequately equipped to provide sufficient support that accounts for the sensitivity of genetic data. This shortcoming can become particularly unsafe when drafting the agreements with the sample providers. This would increase the risk for all the parties involved or pose restrictions that further exacerbate known disparities in research funding and general opportunity between research institutions.

Risks with hesitation to distribute summary statistic data

1. Guaranteed risk to delay our understanding, diagnosis, and treatment of human disease

Our current understanding of the genetics of psychiatric, gastrointestinal, cardiovascular, and autoimmune diseases has been catalyzed by the sharing of summary statistic data to an unimaginable degree. Barriers to entry can be as high as a six month to a two year delay to access data in traditional venues like dbGAP. Given the time constraints of researchers low barrier resources such as ExAC and GnomAD are crucial for advancement of science. These resources have been exemplary in the power to disseminate summary statistic results to the research community. The use of such data has had tremendously positive consequences, and thus far no activity (as far as we are aware) has been harmful to either the PI or the research participants. In the scenario where distribution of summary statistic data is appreciably delayed there is a guaranteed risk to delay our understanding, diagnosis, and treatment of human disease.

2. Increased risk to sampling bias

Hesitation to distribute summary statistic data, we predict, will exacerbate the issues of sampling bias. For convenience most genome-wide association studies will continue being from European populations, as these populations on average have better structured medical records, convenient access to research facilities, and quite liberal laws regarding sharing of health record data for scientific research. This sampling bias will further health disparities and prevent medical advancement from reaching populations which may need it the most.

We see both action and inaction to disseminate as risk introducing activities. If our understanding of human health and disease suffers as a consequence of our inability to come up with low-impedance solutions to address what may be a rare event we will fail the public as health funded researchers. Hence, protection should be placed by NIH to protect PIs, institutions, and research participants as well as ***penalize bad actors that use the data for reidentification.***

Below, we list a set of solutions we recommend to NIH. Our goal in this list is to have very low impedance for researchers in accessing the data while discourage abuse:

- 1) Have low barriers for accessing the summary statistic data:
 - a. Organizational e-mail, and
 - b. Prosecutable commitment to not tease out individuals from the data;

- 2) Inform data contributors that there is a commitment to their privacy;
- 3) Put in place actionable guidelines for security breaches and protections to PIs, institutions, and research participants.
- 4) Encourage legislative protection for research participants.

We see the possibility for medical advancement far outweighing the risks to the individual participants, if and only if legal protections and guidelines are put in place, and we believe that the proposal to add a rapid access mechanism and associated legally binding click-through agreement will address these issues.

Signed by:

Manuel A. Rivas, Armin Pourshafeie, Genevieve L Wojcik and Carlos D. Bustamante

Submission Date: 10/21/2017

Name: Joon Yong An

Name of Organization: University of California, San Francisco

Type of Organization: University

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

I strongly oppose the proposal to lock summary statistics behind a click-through agreement. Limiting data access or adding an extra layer of data access will hamper genetic research and projects that might use such data, cost, and slowing down scientific research.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

There is little risk to misuse of summary statistics. In addition, there is no benefit from a click-through agreement to diminish a risk.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

As we know, the studies showing a risk sensitive to summary statistics were based on a weak model that do not reflect the real world datasets. Homer et al. (2008) provided a model based on homogeneous population, which does not exist in the world. Given the extreme complexity and heterogeneity, there is less risk in population level datasets and no reliable way to re-identify personal information from the dataset.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Submission Date: 11/27/2017

Name: Kenneth Mukamal

Name of Organization: Beth Israel Deaconess

Type of Organization: Healthcare Delivery Organization

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**
- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**
- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

Privacy rules in the US and EU are currently at substantial odds. Multinational studies cannot currently plausibly download genome-wide results if they include EU data. A policy that does not account for nationwide regulations regarding data access will necessarily lead to less generalizable, more insular studies - potentially reducing the value of data-sharing.

Submission Date: 11/28/2017

Name: Mark McGary

Name of Organization: Thyenmed

Type of Organization: Biotech/Pharmaceutical Company

Role: Scientific Researcher

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Subsequent to RSV genome assay, BRCC1 implication was modeled at homology in case. As the case presentation, individual sequence data would be typical, post biopsy, individual case labeling might occur. Any expectation of case privacy, might be compromised in large data where name or case number association could be within data-finding.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Controlled access might be compromised via shared data base ...I.E network. Stand alone or raid *mirror, typical to virtual machine....would seem to solve intrusion, data corruption issues. I have used the current system to good advantage. Why this data may continue to impact research arm, shown in the BRCC1 data, attached below as file. BRCC1 activation to malignancy is a difficult pathogenesis. Precursor causality entire to Beta, within my scope of end use.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

While this is a bit outside of my typical end use, the care and regard among research members, I suspect is very competent. Clearly, the principal investigator would be a key member, as input specific...sensitive data set description.

- 4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies**

V Cell 6, incorporates a web interface where copy and paste features are non-existent. This might be an interesting feature to consider, where any copy to data set is via a key, or adequate log in. I append a recent paper, specific to case tissue with non specific residue as potential patient regard. *Named case (null)
However, there may be tiny sets as site rarity. A difficult privacy outcome, now password secured. It has worked for my end use.

Additional Comment (attachment):

A beta assay of BRCC1 link via Treponema to conservation with estimated precision at RSV homology.

By Mark McGary

The end use of beta molecular toolbar methods via two blast challenge data has created an observational stance of tremode donation within larger viradiae. The high burden of morbidity in neonate at case presentation, as well as the oncogene implication, create a particularly satisfying model to continue within multi species intersection as beta solve. With advance in model at blockade, cleavage, a regard for potential in IND dose chemistry dynamic estimated.

Background.

The clarity of viewpoint, provisioned by the beta solve in pox, lends a capacity not available via typical blast challenge. The pox solve, interesting in the commensal gut tremode donation to pox conservation, when modeled at sample, snout and ventral surface of the reptile, provided a model of cross species jump at predation. (Thyenmed) unpublished.

When inferred as a potential mechanism, inherent to additional viradiae latency as conserved, a method to coherence in beta assay as well as a generalized model as insult modality formed. A high suspicion as to bacterial uptake, to tremode, has also been explored to coherent results in several difficult pathogens. Yersinia pestis, one recent exemplar forged and interestingly, positioned within conservation as continuing to the MDR available mutagenesis predicted to emerge to homo sapiens insult under continuing species pressure via chemotherapeutic dose to primate host population.

Methods:

Traditional NCBI, NIH, Blast, pam 30

Beta fold descriptive blast challenge, Selkov. (Thyenmed) McGary 2014

Authors note:

Regarding the residue complement of 17q.21-31

breast cancer type 1 susceptibility protein isoform 4 [Homo sapiens]

NCBI Reference Sequence: NP_009229.2

Via inquiry, the end use of method (Alan Turing) polymorphism seize, coherent to dualism, is extracted with particular care shown to high mole weight polar individual read, within both fold and down codon estimate to latency as pathogenesis.

*Downcodon 'MDLSALRVEEVQNVINA *from pos 1-

```
Location/Qualifiers
source          1..759
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
```

```

Protein      /chromosome="17"
             /map="17q21.31"
             1..759
             /product="breast cancer type 1 susceptibility protein
             isoform 4"
             /EC_number="2.3.2.27"

```

*Method: Selkov,

Read interior High mole endpoints within a fold character, hence to pam 30 blast challenge specific to Treponema. *Treponema pallidum (taxid:160)

Homo sapiens . Origin via NCBI, NIH

```

ORIGIN
      1 mdlsalrvee vqnvnamgk ilecpiclel ikepvstkc d hifckfcmk llngkkgpsq
     61 cplcknditk rslqestfrs qlveellkii cafqldtgle yansynfakk ennspehikd
    121 evsiiqsmgy rnrakrllqs epenpslqet slsvqlsnlg tvrtlrkqr iqpgktsvyi
    181 elgsdssedt vnkatycevg dqellqitpq gtrdeisl ds akkaacefse tdvtntehhq
    241 psnndlntte kraaerhpek yqgeaasgce setsvsedcs glssqsdlit tqqr dtmqhn
    301 liklqggemae leavleqhg s qpsnsypsii dsalesdlr npeqstseekv ltsqksseyp
    361 isqnppeglsa dkfevsads s tsknkepgve rssp skcpsl ddrwymhscs gslqnrnyps
    421 qeelikvvdv eeqqleesgp hdl tetsylp r qdlegtpyl esgilsf sdd pesdpsedra
    481 pesarvgnip stsalkvpq lkvaesaqsp aaah tdtag ynameesv s r ekpeltaste
    541 rvnkrmsmvv sgltp eefml vykfar khhi t ltnliteet thvvmk tdae fvcertlkyf
    601 lgiaggkwvv syfwvtq sik erkmlnehdf evrgdvvngr nhggpk rare sqdrkifrgl
    661 eiccygpftn mptdqlewmv qlcgasvvke lssftlgtgv hpivvvq pda wtedngfhai
    721 gqnceapvvt rewvldsv al yqcqeldtyl ipqipshy

```

An adequate start position shown via:

CPCLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCK

*subsequent homo sapiens taxid result via: Observing the excellent percent homology.(result just below).

breast cancer type 1 susceptibility protein isoform 4 [Homo sapiens]	144	160	100%	5e-40	100%	NP_009229.2	<input checked="" type="checkbox"/>
<input type="checkbox"/> Select seq gb AAC00049.1	Brca1-delta11b [Homo sapiens]	144	160	100%	5e-40	100%	AAC00049.1
<input checked="" type="checkbox"/> Select seq gb EAW60922.1	breast cancer 1, early onset, isoform CRA_a [Homo sapiens]	144	160	100%	5e-40	100%	EAW60922.1
<input type="checkbox"/> Select seq gb EAW60923.1	breast cancer 1, early onset, isoform CRA_b [Homo sapiens]	144	160	100%	6e-40	100%	EAW60923.1
<input type="checkbox"/> Select seq gb AAU93634.1	breast and ovarian cancer susceptibility protein [Homo sapiens]	144	174	100%	7e-40	100%	AAU93634.1
<input type="checkbox"/> Select seq ref NP_009231.2	breast cancer type 1 susceptibility protein isoform 2 [Homo sapiens]	144	174	100%	8e-40	100%	NP_009231.2
<input type="checkbox"/> Select seq gb ABA29229.1	breast cancer 1 early onset [Homo sapiens]	144	174	100%	8e-40	100%	ABA29229.1
<input type="checkbox"/> Select seq gb EAW60939.1	breast cancer 1, early onset, isoform CRA_I [Homo sapiens]	144	159	100%	8e-40	100%	EAW60939.1
<input type="checkbox"/> Select seq gb ABA29217.1	breast cancer 1 early onset [Homo sapiens]	144	174	100%	8e-40	100%	ABA29217.1
<input type="checkbox"/> Select seq ref NP_009225.1	breast cancer type 1 susceptibility protein isoform 1 [Homo sapiens]	144	174	100%	8e-40	100%	NP_009225.1

BETA Selkov fold origin via:

*BETA FOLD READ AS; CFPKICCFLEILHIDKCEKPTVSVQhV

Blast result at pam 30: *[BLAST](#)® » [blastp suite](#) » RID-1SP7PKA5015 Saved search 11/27/2017

RecName: Full=Uncharacterized protein TP_0177

Sequence ID: [O83207.1](#) Length: 436 Number of Matches: 2

[See 2 more title\(s\)](#)
Related Information

[Identical Proteins](#)-Identical proteins to O83207.1

Range 1: 221 to 233 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
23.1 bits(47)	4.9	Composition-based stats.	9/18(50%)	10/18(55%)	5/18(27%)
Query 25	FLEILHIDKCEKPTVSVQ 42				
	FL+ LHI P VS Q				
Sbjct 221	FLQHLHI-----PSVSAQ 233				

Range 2: 392 to 394 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#) [First Match](#)

Score	Expect	Method	Identities	Positives	Gaps
10.8 bits(18)	37015	Composition-based stats.	2/3(67%)	3/3(100%)	0/3(0%)
Query 31	IDK 33				
	I+K				
Sbjct 392	IEK 394				

Discussion:

Few regards to molecular toolbar advantage Blast Pam 30 output, as perhaps well met as the observational stance of protein read as library at dualism.

Inherent to all species, the regard for host latency to pathogenesis deterministic via dualism library, the authors end use of investigational method shown coherent for difficult pathogen insult potentials.

In the data, positional as study data both to the neonate, as well as the fully developed female homo sapiens case presentation.

```
*RSV, *1..759 /product="breast cancer type 1 susceptibility protein
isoform 4"
/EC_number="2.3.2.27"
```

Respectively. (just above).

Advance to dose build, continuing bio available source read, and tissue based evidence, premised.

McGary

Authors note:

The remarkable incidence at case in neonate, where at 24 months sample, a large swath of neonate exemplar, positive for RSV,
As well as the past and present morbidity and mortality of the Yersinia, treponema, demonstrated at homology...continue to occupy features of
research rm composite now investigational to 2030 goals.

References:

NCBI, NIH, Blast (Pam 30) *Saved search available upon request

NCBI, NIH, Cobalt

Thyenmed (Duns) 079277748

Dr. Emma Saavedra

*Research, reviews.

Name: Liz Malerba

Name of Organization: United South and Eastern Tribes Sovereignty

Protection Fund **Type of Organization:** Tribal

Role: Government Official

Information Requested

- 1. Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

Regarding data obtained from the citizens of Tribal Nations, any access mechanisms must have informed consent not only among the study participants, but also designated officials within the Tribal Nation. This consent mechanism varies from Tribal Nation to Tribal Nation and may take the form of Tribal Nation Council resolutions, signed MOU's with the designated Tribal leader, etc. Before any research is conducted upon American Indians and Alaska Natives (AI/AN) and/or in Indian Country, researchers must provide evidence of Tribal Nation informed consent, which must address issues surrounding data use, data ownership, publication permissions, and specimen policies, among other issues. Historically, Tribal Nation consent has not been sought or obtained, and because of this, there are many documented cases of the exploitation of Tribal Nation data causing harm to Tribal Nation communities. Therefore, it is USET SPF's position that:

A. Any Tribal Nation data currently in the database be removed unless Tribal Nation consent is received.

B. All future research studies occurring in Indian Country require Tribal Nation consent, which addresses, at a minimum, the issues outlined above.

- 2. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity**

No Tribal Nation data should be included in any level of access without explicit Tribal Nation consent. Even if this data is publicly available elsewhere, NIH should honor its federal trust obligation as well as the government to government relationship between US and Tribal Nations, and should thus require all submitting institutions to submit evidence of Tribal Nation consent.

- 3. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive**

It is not appropriate for the submitting institutions alone to decide which Tribal Nation datasets should be considered ‘sensitive.’ There must be written documentation from the Tribal Nation indicating its willingness to participate in the database.

4. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies

USET SPF asserts that shared power and decision making between the researcher and the Tribal Nation, utilizing the principals of Participatory Research (PR), should become the standard, and indeed be required, by NIH for all studies conducted in Indian Country. It is thus USET SPF’s position that the questions above be reframed to be culturally relevant to Indian Country and include language that reflects PR. As an agency of the federal government, NIH has a duty to protect the data of Tribal Nations and ensure data management policies are crafted in collaboration with Tribal governments.

See attached document for further information.

Additional Comment (attachment):

December 12, 2017

Office of Science Policy
National Institutes of Health
6705 Rockledge Drive
Suite 750
Bethesda, MD 20892

Re: USET SPF Comments on the NIH Proposal to Update Data Management of Genomics Summary Results

The United South and Eastern Tribes Sovereignty Protection Fund (USET SPF) is pleased to offer the following comments on the National Institutes of Health (NIH) Proposal to Update Data Management of Genomics Summary Results. USET SPF recognizes the importance of Genomic Data, as we believe this information will be the foundation for future scientific innovation. However, it is critically important to recognize the historic relationship between scientific study and Tribal Nations, where researchers committed ethical violations against our communities and our people. American Indians and Alaska Natives (AI/AN) and Tribal communities have experienced negative impacts from the use of genomic data (*Arizona Board of Regents v. Havasupai Tribe*) without Tribal Nation informed consent. To ensure the privacy of Tribal Nation communities, as well as AI/AN individuals, USET SPF urges NIH to consult with Tribal Nations regarding its research and data management policies and offers the following recommendations in response to NIH's request for comment.

USET SPF is a non-profit, inter-tribal organization representing 27 federally recognized Tribal Nations from Texas across to Florida and up to Maine.¹ Both individually, as well as collectively through USET SPF, our member Tribal Nations work to improve health care services for American Indians. Our member Tribal Nations operate in the Nashville Area of the IHS, which contains 36 IHS and tribal health care facilities. Our citizens receive health care services both directly at IHS facilities, as well as in Tribally Operated facilities operated under contracts with IHS pursuant to the Indian Self-Determination and Education Assistance Act (ISDEAA), P.L. 93-638.

- i. **Risks and benefits of providing broad access to genomic summary results from most studies in NIH-designated data repositories utilizing the proposed rapid access mechanism and associated click-through agreement. Risks and benefits may relate to participant protection issues and/or scientific opportunity.**

Regarding data obtained from the citizens of Tribal Nations, any access mechanisms must have informed consent not only among the study participants, but also designated officials within the Tribal Nation. This consent mechanism varies from Tribal Nation to Tribal Nation and may take the form of Tribal Nation Council resolutions, signed MOU's with the designated Tribal leader, etc. Before any research is conducted upon American Indians

¹ USET SPF member Tribal Nations include: Alabama-Coushatta Tribe of Texas (TX), Aroostook Band of Micmac Indians (ME), Catawba Indian Nation (SC), Cayuga Nation (NY), Chitimacha Tribe of Louisiana (LA), Coushatta Tribe of Louisiana (LA), Eastern Band of Cherokee Indians (NC), Houlton Band of Maliseet Indians (ME), Jena Band of Choctaw Indians (LA), Mashantucket Pequot Indian Tribe (CT), Mashpee Wampanoag Tribe (MA), Miccosukee Tribe of Indians of Florida (FL), Mississippi Band of Choctaw Indians (MS), Mohegan Tribe of Indians of Connecticut (CT), Narragansett Indian Tribe (RI), Oneida Indian Nation (NY), Passamaquoddy Tribe at Indian Township (ME), Passamaquoddy Tribe at Pleasant Point (ME), Pamunkey Indian Tribe (VA), Penobscot Indian Nation (ME), Poarch Band of Creek Indians (AL), Saint Regis Mohawk Tribe (NY), Seminole Tribe of Florida (FL), Seneca Nation of Indians (NY), Shinnecock Indian Nation (NY), Tunica-Biloxi Tribe of Louisiana (LA), and the Wampanoag Tribe of Gay Head (Aquinnah) (MA).

and Alaska Natives (AI/AN) and/or in Indian Country, researchers must provide evidence of Tribal Nation informed consent, which must address issues surrounding data use, data ownership, publication permissions, and specimen policies, among other issues. Historically, Tribal Nation consent has not been sought or obtained, and because of this, there are many documented cases of the exploitation of Tribal Nation data causing harm to Tribal Nation communities. Therefore, it is USET SPF's position that:

- A. Any Tribal Nation data currently in the database be removed unless Tribal Nation consent is received.
- B. All future research studies occurring in Indian Country require Tribal Nation consent, which addresses, at a minimum, the issues outlined above.

ii. Risks and benefits of maintaining genomic summary results from studies designated by the submitting institution to include sensitive information in controlled access. Risks and benefits may relate to participant protection issues and/or scientific opportunity

No Tribal Nation data should be included in any level of access without explicit Tribal Nation consent. Even if this data is publicly available elsewhere, NIH should honor its federal trust obligation as well as the government to government relationship between US and Tribal Nations, and should thus require all submitting institutions to submit evidence of Tribal Nation consent.

iii. Appropriateness of the proposal for institutions submitting study data under the NIH GDS Policy to indicate which datasets should be designated as sensitive.

It is not appropriate for the submitting institutions alone to decide which Tribal Nation datasets should be considered 'sensitive.' There must be written documentation from the Tribal Nation indicating its willingness to participate in the database.

iv. General comments on any other topic relevant to unrestricted, rapid, or controlled-access to genomic summary results from NIH-funded studies.

USET SPF asserts that shared power and decision making between the researcher and the Tribal Nation, utilizing the principals of Participatory Research (PR), should become the standard, and indeed be required, by NIH for all studies conducted in Indian Country. It is thus USET SPF's position that the questions above be reframed to be culturally relevant to Indian Country and include language that reflects PR. As an agency of the federal government, NIH has a duty to protect the data of Tribal Nations and ensure data management policies are crafted in collaboration with Tribal governments.

Conclusion

USET SPF appreciates this opportunity to provide comments on the NIH proposal to Update Data Management of Genomics Summary Results. Because this proposal has significant implications for Tribal governments and their citizens, we urge NIH to seek formal Tribal Consultation on this specific issue, as well as the larger issue of protection of Tribal Nation data. Should you have any questions or require additional information, please do not hesitate to contact Ms. Liz Malerba, USET SPF Director of Policy and Legislative Affairs, at (202) 624-3550 or by e-mail at lmalerba@usetinc.org.

Sincerely,

Kirk Francis
President

Kitcki A. Carroll
Executive Director