MEETING SUMMARY

Innovation Brainstorm:

Transforming Discovery Into Impact

May 4-6, 2011

Table of Contents

Executive Summary 4

Beyond GWAS 5

Microbiome Studies 5

Group Effects 6

Artificial Organs as Tools for Translation 6

Proteomics and Therapeutics Development 7

Single-Cell Analysis 7

NIH Award Strategies 8

Computational Challenges 8

Molecular Classification of Disease 9

Day 1: Introduction and Welcome 10

History and Overview of the NIH Common Fund 10

Day 2: "Fishbowl" Discussion of Pre-Selected Papers 14

I. Beyond Genome-Wide Association Studies (GWAS) 14

II. Microbiome and Global Health 17

III. Group Effects (combinations of infectious and environmental agents, drugs) 19

IV. Artificial Organs as Tools for Translation 21

V. Proteomics and Therapeutics Development 23

VI. Single Cell Analysis 25

VII. NIH Award Strategies 27

VIII. Other 29

Day 3: Group Summaries and Recommendations 33

I. Beyond GWAS 33

II. Microbiome: Part 2 34

III. Group Effects 36

IV. Artificial Organs as Tools for Translation 36
V. Targeting the Dynamic Complexome 38
Va. Bringing Difficult Structures into Reach 38
VI. Single-Cell Analysis 39
VIa. Cross-Cutting Issues in Computation and Informatics 40
VII. NIH Award Strategies 41
VIII. Molecular Classification of Disease 42
Closing Remarks 43
APPENDIX A: Selected CF Programs 45
APPENDIX B: INVITED PARTICIPANTS 48
APPENDIX C: NIH PARTICIPANTS 51

Innovation Brainstorm: Transforming Discovery into Impact

May 4-6, 2011

Bolger Center, Potomac, Maryland

Executive Summary

The National Institutes of Health (NIH) held the 2011 Innovation Brainstorm meeting to identify areas of emerging scientific opportunities in which strategic investment by the NIH Common Fund (CF) could accelerate progress. Participants were selected to draw from the creativity and interdisciplinary expertise of early-career investigators who have been recognized through various award competitions as broad-thinking emerging leaders in their fields.

Strategic planning for the CF is undertaken regularly to identify cross-cutting research areas that have high potential for transformative impact. After external and internal input, program areas are refined into specific initiatives that are informed by NIH portfolio analysis and input from expert panels. In past years, the NIH has sought external input from senior thought leaders to identify scientific needs and opportunities for CF strategic planning. This year, Innovation Brainstorm was a new approach to solicit fresh ideas and new perspectives about exciting, high-impact research opportunities where funding through the CF may have an important impact.

In advance of the Innovation Brainstorm meeting, participants were asked to identify an exceptionally innovative, groundbreaking, or cross-cutting paper from the past year that illustrates a high-risk or emerging area of science with great potential to be transformative. All papers were made available to the entire group, and participants discussed the selected papers via an online forum. In addition to identifying exciting and emerging areas of science, participants also generated ideas for how strategic investments could accelerate the potential impact of the science.

The online discussion of the scientific papers by the participants led to the following topic areas selected for further discussion at the Innovation Brainstorm meeting.

- Beyond Genome-Wide Association Studies (GWAS)
- Microbiome Studies
- Group Effects
- Artificial Organs as Tools for Translation
- Proteomics and Therapeutics Development
- Single-Cell Analysis
- NIH Award Strategies

Each topic was discussed through a “fishbowl” conversation strategy to encourage dynamic, yet focused, dialogue with all meeting participants. Through discussion, two additional topics emerged that cut across many of these areas and were also presented during the meeting.

- Computational Biology and Informatics

- Molecular Classification of Disease

The next steps in CF strategic planning will include opportunities for IC Directors to also propose areas of scientific needs and opportunities. These topic areas will be refined, prioritized by senior NIH leadership, and assigned to Working Groups for analysis and development into specific program initiatives that can be supported through the Common Fund beginning in FY 2013.

The topics discussed during the Innovation Brainstorm meeting are summarized below with a list of specific recommendations made by the meeting participants.

Beyond GWAS

Although GWAS have uncovered many genetic loci for a range of conditions and diseases, a major challenge is translating this knowledge into functional insights. There also remains a significant computational challenge that calls for innovative algorithms to analyze genotypic and phenotypic data. Improved collection, standardization, and integration of data will inform the study of pathways common to various diseases.

Recommended CF investments in this area include:

- Establish a human "phe-ge" project — a very large-scale effort to create a "national cohort" of people (DNA plus phenotypic data) for discovery research in health and disease.
- Establish a functional genome project that leverages functional information to find causal variants — employing ENCODE, epigenomics, and functional genomics strategies.
- Establish a "multidimensional analyses for genomic studies" project that provides context for genomic data by accessing environmental measures, incorporating population and family structure, and including epigenetic context.

Potential impact: Moving GWAS beyond its current capability offers faster movement from association to function, which will likely accelerate discovery for multiple traits.

Microbiome Studies

While the Human Microbiome Project has provided important funding for large-scale sequencing projects to determine the composition of bacteria in various tissues, future funding on mechanistic studies will be crucial for understanding how commensal bacteria can confer protection against disease. Advanced methods are needed to deconvolute data representing complex mixtures of species (host, commensals, and pathogens) as well as their interactions.

Recommended CF investments in this area include:

- Develop improved model systems of co-infection and organ-host systems.
- Support projects that explore the mechanisms through which the microbiome influences health and go beyond correlation.
- Support microbiome-wide association studies (MWAS).

- Develop computational tools that will allow deconvolution of complex data sets.
- Support the functional analysis of small molecules derived from microorganisms.
- Study the systemic impact of the gut microbiome on other organs.
- Examine mother-child microbiome effects.
- Any NIH-sponsored efforts in this area should be multidisciplinary to realize the fullest impact of potential.

Potential impact: Systems biology-derived "designer" probiotics may offer an inexpensive, holistic approach to disease prevention and treatment.

Group Effects

Although the impact of infectious agents, environmental exposures, drugs, and other stressors are generally studied one at a time, we often experience them in combination. We need systematic unbiased screens for studying how multiple factors (e.g. microbiological, chemical, lifestyle and dietary exposures) interact to contribute to susceptibility to disease, disease progression, and treatment outcomes.

Recommended CF investments in this area include:

- Develop quick and cheap technologies to measure the following systematically and in a natural environment:
    - Multiple environmental and dietary exposures
    - Multiple infections
    - Multi-drug effects
    - Combinations of any of the above
- Support the development of bioinformatic techniques for analysis of large clinical datasets
- Create and curate a database that lists and characterizes model systems suitable for studying multiple factors that may contribute to disease.

Potential impact: Progress in this area will advance understanding of mechanisms through which combinations of factors influence health in ways that are not simply additive. This will lead to the identification of preventive strategies that take these factors into account for a personalized medicine approach.

Artificial Organs as Tools for Translation

Tissue engineering is an attractive area for many reasons, not the least of which is its near-term clinical applicability for accelerating drug development through the provision of reliable toxicological testing models. There is considerable enthusiasm for simulating the development of major organ systems using human stem cells (iPS cells in particular) and/or combining them with bioengineering approaches.

Recommended CF investments in this area include:

- Develop and support technology transfer mechanisms that accelerate tool- and instrumentation sharing.
- Integrate developmental biology, organ-quality assays and bioengineering.
- Develop iPS cell drug-screening platforms to replace traditional tests.
- Establish an NIH-funded "Jackson Lab" for organ-mimics.
- Any NIH-sponsored efforts in this area should be multidisciplinary to realize the fullest impact of potential.

Potential impact: Progress in this area may contribute to a paradigm shift in conducting clinical trials, substantial advances in developmental biology understanding, the next-generation "immunoassay," and ultimately the development and use of artificial organs.

Proteomics and Therapeutics Development

Despite ongoing CF efforts to develop new technologies and a better understanding of the structural biology of membrane proteins, many other proteins and protein complexes of intense biological interest remain intractable to structural biological investigation. Numerous proteins contribute to more than one protein complex and many of these are transient or dynamic in nature with diverse functions. The localization, interactions, posttranslational modifications and relative expression levels contribute to both protein and protein complex function in homeostasis and disease. Experimental mapping of the dynamic "complexome" in normal and disease states would significantly accelerate translational research in the uncharted multi-dimensional space between genomics and personalized drug discovery.

Recommended CF investments in this area include:

- Develop new technologies to characterize functionally distinct, temporally dynamic protein complexes, including those found in diseases.
- Develop computational tools and algorithms that allow predictive models for protein-protein and protein-drug interactions to be established and tested.
- Develop effective drugs that target complexes through a) rational design and b) screening of protein/small molecule interactions.
- Sponsor workshops to define and understand the current limits of emerging structural biology technologies, prioritize those for development, and improve access to and understanding of these techniques.

Potential impact: Understanding protein interactions and activities in a spatial and temporal context provides the opportunity to design "smarter," targeted therapeutic interventions that acknowledge and address the complexity of protein interactions and dynamics. Mapping the complexome may have as much or more potential for distinguishing disease states as does mapping the genome.

Single-Cell Analysis

Single-cell analysis is a research endeavor that bridges most biomedical fields and interests, and its value extends beyond fundamental research or technology development. One main reason for the broad relevance of single-cell approaches is because most studies of cells and their behavior have been

performed on populations of cells that we now know are, in most cases, highly heterogeneous. Thus, many standing conclusions in biomedicine have been based upon averages of ensembles of cells.

Recommended CF investments in this area include:

- Develop new technologies to map a single cell's epigenome, proteome, and metabolome.
- Extend recent proof-of-principle work in single cell genomics and transcriptomics that is highly innovative, but low-throughput and far from practice.
- Employ approaches that capture living (or recently living) cells *in vivo* without the need for overexpression or artificial constructs.

Potential impact: The ultimate goal of this field of research is the ability to monitor (and ultimately, manipulate) *in situ* all of the parameters of single-cell behavior within populations and subpopulations.

NIH Award Strategies

There continues to be an unmet need to bring together innovative researchers from distinct fields. Many view the most successful cross-disciplinary collaborations as being self-assembled and not too large. A key bottleneck appears to be finding collaborators, since investigators within different disciplines attend different sets of meetings, apply for different funding, read different literature, and speak different "languages."

Recommended CF investments include:

- Create more opportunities for multidisciplinary teams to form.
- Create pilot Ph.D. programs in emerging, cross-disciplinary areas.
- Target postdoctoral fellowship awards to physical scientists and engineers.
- Facilitate Ph.D. and M.D. interactions.
- Create mechanisms for small-team proposals.
- Fund an intermediate stage of funding that validates university or donor-sponsored initiatives.

Computational Challenges

One common thread of nearly all the topics discussed at the 2011 Innovation Brainstorm is data overload. In particular, there is an urgent need for integration of data sets, approaches — as well as of inquiry that addresses multiple states of health and disease. Improved data sharing, as well as access to secondary data sets, is paramount to progress.

Recommended CF investments include:

- Create innovation centers.
- Lower the entry barrier for quantitative scientists.
- Establish and present "prediction challenge" data sets for multidisciplinary teams to solve.
- Develop and/or host a software commons for computational biologists and enhance the usability of web-based computational tools for biologists.

- Create an Office of Cyberinfrastructure within the CF.

Potential impact: Developing and sharing broad-based computational tools and making them freely available to the scientific community has the potential to vastly increase the interoperability of data sets currently being generated in numerous studies.

Molecular Classification of Disease

Currently, “clinical syndromes” are usually used to classify disease. New approaches are needed to classify patients and disease states via molecular determinants.

Recommended CF investments in this area include:

- Support expression analysis of patient samples.
- Support epigenetic analyses of distinct cell types.
- Support classification and measurement of behavioral symptoms.
- Integrate different phenotypes.
- Develop methods to measure the response to various perturbations.

Potential impact: Removing barriers to translation has obvious benefit across the board for the diagnosis and treatment of all diseases. Progress in this area would catalyze the transition from one-size-fits-all medicine to personalized medicine.

Day 1: Introduction and Welcome

NIH Director Dr. Francis Collins welcomed the group of scientists representing various interests within the NIH mission to the 2011 Innovation Brainstorm. He noted that this meeting is the current iteration of a series of information-gathering efforts that inform the development of initiatives for the NIH Common Fund[1] (CF). As with previous meetings of this sort, NIH staff welcome an informal and candid discussion of current needs and opportunities in biomedicine, toward identifying ideas ripe for near-term investment that are likely to have significant impact for science and public health.

History and Overview of the NIH Common Fund

In 2003, under the leadership of then-NIH Director Dr. Elias Zerhouni, the NIH recognized the need to establish the means to prepare for a changing biomedical landscape, and in particular that the structure of the NIH was not ideal to achieve goals on a short timetable. Moreover, there existed no formal processes to support cross-cutting science that traversed Institute and Center (IC) mission boundaries. This situation presented a need to i) accelerate the pace of discoveries in the life sciences; ii) enable deeper understanding of the pathobiology of disease prior to irreversible damage; iii) translate research more rapidly from laboratories to patients and back; and iv) explore novel biomedical strategies orders of magnitude more effective than current ones. A series of internal and external discussions led to the development of the NIH Roadmap for Medical Research. Its goals were to accelerate basic research discoveries and speed translation of those discoveries into clinical practice, as well as to explicitly address roadblocks that slow the pace of medical research toward improving the health of the American people. The NIH Roadmap was codified in 2006, with passage of the NIH Reform Act of 2006 that also created the Division of Program Coordination, Planning, and Strategic Initiatives (DPCPSI) within the NIH Office of the Director, and the NIH CF to facilitate trans-NIH research, which is intended to provide the NIH with an intellectual venture capital space for promoting innovation.

In addition to the CF, DPCPSI also houses other trans-NIH programs, including the Office of AIDS Research, the Office of Research on Women's Health, the Office on Disease Prevention, and the Office of Behavioral and Social Sciences Research. The Office of Strategic Coordination within DPCPSI manages the CF as an incubator for time-limited programs that address specific challenges and catalyze IC-funded work. By law, the CF is a set, protected proportion of the NIH enacted budget; thus the CF budget rises and falls in concert with the agency's Congressional appropriation and cannot shrink below this proportion (in FY 2011, $543 million, which is 1.7 percent of the NIH budget). CF criteria include that projects must be:

- Transformative: Must have high potential to dramatically affect biomedical and/or behavioral research over the next decade
- Synergistic: Outcomes must synergistically promote and advance individual missions of NIH ICs to benefit health

[1] http://commonfund.nih.gov/

- Cross-Cutting: Program areas must cut across missions of multiple NIH ICs and be relevant to multiple diseases or conditions
- Broad Benefit: Must be something no other entity is likely or able to do, and research results must benefit public health

As the funding of original CF projects ends in coming years, more funding will open up for new investments (~ FY 2013). Projects are expected to move out of the CF space within a specified time frame, at which point they either expire or are adopted by specific ICs. There is some debate, however, about retaining within the CF certain initiatives such as the High-Risk, High-Reward Research (HRHR) initiatives that support exceptionally creative scientists through the Pioneer, New Investigator, and other programs that currently comprise approximately one-third of the CF budget. While individual projects will turn over within the 5-10 year timeframe, the initiatives themselves may continue to be part of the CF to provide unique opportunities for innovation to the entire community. Alternatively, new funding mechanisms may be tested with the HRHR Program to explore new ways through which the NIH can foster innovation and creativity.

The CF also provides the NIH Director an opportunity to respond quickly to public health threats or other rapidly emerging opportunities for transformative investment. One recent example is the funding of a health-monitoring study to follow workers involved in the Gulf oil spill cleanup following the Deepwater Horizon explosion and its consequences on the environment and on public health. Currently under conceptual and logistical development is a unique, intramural component of the CF: the Center for Regenerative Medicine, which aims to provide resources as well as facilitate translation and clinical study using various stem cell models.

CF programs span a range of interest areas and mechanisms, but all aim to be catalytic, as i) tools, infrastructure and data streams to establish new areas of study[2]; ii) technologies and approaches to overcome barriers within a field[3]; or iii) new approaches to foster innovation and creativity[4]. Dr. Collins provided descriptions and contextual information about several CF projects so that the meeting participants would have a firmer understanding of new programs that might be developed as a result of this meeting. A summary of these programs is offered as Appendix A of this report.

Dr. Collins then explained that priority-setting for the CF occurs through a two-phase process, on an approximately yearly basis. First, input is gathered through a targeted meeting (like the Innovation Brainstorm) as well as from NIH staff and leadership. Second, NIH staff refine the list of identified programs through portfolio analyses and discussion. Senior NIH leadership review the plans, and the NIH Director makes the final decision about which of the programs will receive Common Fund support.

[2] Examples include: Molecular Libraries and Imaging, Human Microbiome Project, Genotype-Tissue Expression Resources (GTEx), Protein Capture Reagents

[3] Examples include: Structural Biology; Technology Centers for Networks and Pathways

[4] Examples include: Interdisciplinary Research, High-Risk High-Reward (Pioneer Awards, New Innovator Awards, Transformative R01 (TR01) awards, Early Independence Awards)

The NIH acts strategically to fund CF proposals, primarily through testing new approaches to speed scientific progress in tackling emerging opportunities in biomedical research that have the potential for high impact. Projects at or near the inflection point of a graph plotting scientific progress over time are the most ripe for investment of this type.

Questions and Discussion

A significant issue relating to all currently funded CF projects, and most new candidate programs, is data volume. The National Centers for Biomedical Computing[5] were put into place in 2004, intending to address the issue of creating shared resources and tools for data management and analysis. Projects funded to date have been successful in their own right, but they have been more focused in research goals, and only limited in their impact on interoperability. Many of these projects have migrated to individual ICs for continued funding. There is concern that greater emphasis is needed on cross-cutting tools and resources, as well as funds for curation, integration, and storage for the longer term. Needs extend beyond software tools, and include innovation in computational biology, machine learning, and other techniques and approaches common to the worlds of physics, engineering, and computer science.

A key bottleneck in moving biology and the quantitative worlds closer to each other is a dearth of "bilingual" scientists that can recognize/identify problems of significant biomedical importance but are also conversant in mathematics, advanced statistics, computer science, and physics. The problem of integrating biology with the quantitative sciences goes well beyond finding "matchmaking" expertise to solve each other's problems: What may be the most needed is dual training during college and graduate school. Interdisciplinary cross-training with co-mentoring may be an exemplary approach toward growing this segment of the biomedical workforce. Another roadblock beyond the purview of the CF is peer review: Quantitative scientists' biology-geared applications often fare poorly in review. Exacerbating the problem is that quantitative reviewers appear to be in ever short supply.

All told, it appears that although data deluge is itself a problem to be reconciled, a greater urgency is the need to integrate various data formats — a goal that calls for much more synchrony in the standardization of data collection and significant funds. Another consideration is whether all the data currently being collected is necessary to answer specific research questions, or whether it is being collected because technology enables it to be. Much of the innovation in this realm is likely to come from individual investigator creativity and ingenuity in creating smart tools to mine and integrate data, or in biological engineering approaching that build living systems from scratch, then perturb their function to understand the rules of biology better. There is some consensus that data integration is not field-specific, and thus suitable for the cross-cutting nature of CF investment. For example, broad-based annotation of genomic data, signaling circuitry, pathways common to many diseases (e.g., inflammation), and other parameters characterizing biological data, may be of significant use across ICs and diseases. Parallel efforts must assure that clinicians acquire the necessary knowledge and tools to understand and use results acquired through genomic studies and other high-throughput approaches, and that ethical, legal, and social (ELSI) concerns are addressed. Much needed are methods to catalyze a

[5] http://www.ncbcs.org/

more seamless acquisition of clinically relevant data from electronic medical records (EMR) and clinical trial data. The currently operational HMO Collaboratory[6] is a "laboratory" to begin to address some of these issues on a relatively large scale.

The CF may be an important vehicle for expanding "failure analyses," to capture data and results from previous experiments that did not support the original hypothesis under study. The newly created National Center for Advancing Translational Sciences (NCATS) aims to address this issue by re-analyzing failed drug molecules and cross-testing molecules with new targets. NCATS also aims to help bridge the pre-clinical research "valley of death," in which many "hits" fail. Another goal of NCATS is to support process innovation in model systems and regulatory science. Making pre-clinical testing more predicative might be accomplished through iPS cell systems or organoids that can effectively screen pharmacokinetic and pharmacodynamic parameters. It is not yet clear whether these sorts of investments will fall within the CF or within NCATS.

Currently, one-third of the CF expenditure goes toward programs that aim to harness the creativity of exceptional minds. Biomedical research is a human endeavor, and the NIH remains concerned that systemic effects that are delaying the time to career independence in the life sciences are quashing creativity. Thus, recently, a new advisory group to the NIH Director was established: the group's charge is to look systematically at the growth and health of the biomedical research workforce. This effort is being conducted to better understand supply and demand issues.

[6] http://commonfund.nih.gov/hmocollaboratory/

Day 2: "Fishbowl" Discussion of Pre-Selected Papers

The 2011 Innovation Brainstorm meeting represents a new approach to identifying topic areas that may be ripe for CF investment. Attendees were each asked to identify, describe, and defend — in advance, and through an online forum — the most exciting paper they had read in the past year. They were asked to consider the potential impact of the paper on a broad spectrum of health research and to identify ways in which CF investment might allow the impact to be realized sooner rather than later. The online discussion included an opportunity for each participant to nominate a subset of papers for further discussion at the meeting itself. Meeting sessions were then formed as a basis of this online discussion.

Discussion at the meeting then occurred through a "fishbowl" conversation[7]: a dialogue that can be used with large groups to encourage active but controlled participation by all who are present. During the Innovation Brainstorm meeting, six chairs were arranged in an inner circle (the fishbowl), with the remaining chairs arranged in concentric circles outside it. Five participants[8] were pre-selected to fill the fishbowl, while the rest of the group sat on the chairs outside the fishbowl. One chair was left empty. At any point in the discussion, a participant from the outer circles could join the conversation by occupying the empty seat; this would prompt a member of the fishbowl to exit, thereby freeing a seat for another participant from the outer circles.

The initial participants inside the fishbowl for each discussion included those participants who had nominated the papers that served as the basis for that session plus a moderator, who was an NIH expert in that area. Each participant who selected one of the papers being discussed gave a 3-5 minute synopsis of the paper's major findings and impact on a given field. All members of the fishbowl then discussed suggestions for potential NIH CF investment, with everyone in the room rotating in and out of the fishbowl as desired. When time ran out, the fishbowl was closed and the moderator summarized the discussion.

A contextual summary of the in-person discussion of each topic appears below. The overall purpose of the discussion was to identify barriers and potential impact, as well as to devise strategies for reducing the time to impact.

I. Beyond Genome-Wide Association Studies (GWAS)

In this session, discussion leaders presented and discussed four papers.

1. Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes.[9]

Discussion leader Dr. Jose Florez, Massachusetts General Hospital, provided the following summary:

[7] http://en.wikipedia.org/wiki/Fishbowl_(conversation)

[8] Participants who nominated the papers upon which the discussion topic is based; Other participants who expressed interest in those papers; A moderator from the NIH (expert in the field)

[9] Zhou K, Bellenguez C, Spencer CC, Bennett AJ, et al. Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes. *Nat Genet*. 2011;43:117-20.

This landmark study is one of few GWAS that evaluates drug response: in particular, to metformin, the first-line agent for diabetes therapy. Though the discovery cohort is modestly sized, it is powered to detect relatively strong effects conferred by common variants, of a magnitude comparable to that seen in other complex traits. Genotyping of the discovery cohort is high-quality, with appropriate quality control thresholds and filters. The authors provide independent evidence of replication in two additional cohorts, including a prospective clinical trial. Within each cohort, the results are internally consistent in two complementary analytical methods (linear and logistic regression) and survive sensitivity analyses (when restricted to monotherapy). The authors provide persuasive evidence that the association is due to metformin, and not due to hemoglobin A1c itself or the disease process. In summary, the authors go beyond genetic association and conduct functional work to identify the causal gene.

2. Analysis of genetic inheritance in a family quartet by whole-genome sequencing.[10]

Discussion leader Dr. Erin Clark, University of Utah Health Sciences, provided the following summary:

GWAS have contributed to our understanding of the genetic influences on complex diseases. However, the 'common disease-common variant' genetic hypothesis has not been substantiated for many disorders of complex inheritance, including preterm birth and hypertension — diseases with a tremendous and growing public health impact. While some GWAS have been valuable in identifying unlikely suspects of disease, most have failed to identify significant genomic variants. For even heavily studied diseases, the variants found to date typically explain only a fraction of the heritable variance in disease risk. New methodologies to assess more rare genetic variants are therefore necessary to further explore the genetic contribution to complex diseases. Whole-genome and exome sequencing methods have the unique ability to identify the effects (potentially large) of rare alleles in small pedigrees. This approach can complement the power of GWAS in identifying weak effects of common alleles in larger populations. Whole-genome sequencing methods are particularly intriguing since an unknown fraction of important phenotypes in humans are encoded by non-exonic variants that can only be identified by this methodology. This research demonstrates the power of family whole-genome analysis to rapidly identify rare disease-gene candidates in a Mendelian disorder. Perhaps the most exciting observation is that whole-genome sequencing has the potential to locate causative genes in complex diseases for which traditional methods have been uninformative.

3. Genetic variants near TIMP3 and high-density lipoprotein-associated loci influence susceptibility to age-related macular degeneration.[11]

Discussion leader Dr. Hui Sun, University of California, Los Angeles, provided the following summary:

This paper represents the culmination of years of work on the genetic basis of age-related macular degeneration (AMD), the leading cause of irreversible blindness in the United States. This work is the

[10] Roach JC, Glusman G, Smit AF, Huff CD, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science*. 2010;328:636-9.
[11] Chen W, Stambolian D, Edwards AO, Branham KE, et al. Genetic variants near TIMP3 and high-density lipoprotein-associated loci influence susceptibility to age-related macular degeneration. *Proc Natl Acad Sci U S A*. 2010;107:7401-6.

largest scale GWAS of AMD and provides a high-resolution map of major loci contributing to AMD in the human genome. No one could have predicted before 2005 (when the fist GWAS was done) that five of the top six genes associated with AMD in the human genome would be involved with innate immunity. Although this result is a confirmation of earlier studies from other groups, this paper offers a beautiful big picture that puts all previous findings in perspective. Discovery of these major loci associated with AMD provides the unprecedented opportunity to understand AMD etiology and to develop more effective therapies that target the underlying causes. However, AMD genetics illustrate the challenges to fully realize the therapeutic potential of GWAS. Despite a large number of papers published on this subject, the true, causal SNPs for the two most important AMD loci have not been identified, as discussed in this paper. Even the identities of the genes (not just the SNPs) have been questioned by newer genetic studies. These ambiguities have substantially hindered further progress, because there are still many possible pathogenic pathways if the causal SNPs are not identified. Elucidating the specific triggers and pathways is critical in developing the most effective AMD treatment, because general inhibition of innate immunity is unlikely to be the best option.

4. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations.[12]

Discussion leader Dr. Bradley Malin, Vanderbilt University Medical Center, provided the following summary:

Until recently, EMR were deemed to be too noisy to perform meaningful clinical research. However, this paper demonstrates that not only can we use EMR to confirm genotype-clinical phenotype relationships, we can discover potentially new clinical associations. In particular, this project flips the concept of GWAS on its head. Rather than compute the relationship between a single clinical concept and a large number of genetic regions, this work computes the relationship between a single genetic region and all clinical concepts; that is, a phenome-wide association study (PheWAS). In doing so, this work illustrates that clinical phenomena in a common pathway can be discovered (or at least a hypothesis for their relation can be generated). The challenge of this work is in the repurposing of the EMR system. The authors demonstrate that the construction of a positive (or negative) phenotype from a medical record cannot be performed haphazardly. One cannot simply test the correlation between a genotype and diagnosis codes in the EMR system. Rather, it is critical to use multiple modes of evidence (e.g., multiple occurrence of a clinical diagnosis, confirmation of diagnosis in diagnosis codes and laboratory reports, etc.). This research demonstrates that "on the shelf" EMR systems can support novel clinical-genomic investigations. However, there are noise and small sample statistics at play for certain phenotypes. Moreover, it should be recognized that the approach described in this paper is relatively specialized to a particular EMR system. As such, it is critical to develop computational tools and critical infrastructure to realize PheWAS on a broad scale.

Session Discussion

[12] Denny JC, Ritchie MD, Basford MA, Pulley JM, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*. 2010;26:1205-10.

GWAS is a time-tested tool for investigating the genetic basis of disease; however, the study of complex traits and multivariate diseases require huge sample sizes to provide adequate statistical power. New methods are needed for robust quality control applied to emerging sequence data, a repository for analysts to test those tools, adequate storage capacity, and appropriate ethical safeguards. There is broad consensus that, at some level, there exists a significant computational challenge that calls for innovative algorithms for analyses.

Also needed are multi-tiered approaches for integrating data. In particular, there is an urgent need to translate GWAS findings for clinical relevance and potential use. In the case of AMD, for example — one of the first GWAS-disease "success" stories — the outcome has been less clear than originally imagined. This is primarily due to the fact that although the major genetic influence to this disorder has been identified, it is still not enough to affect disease treatment or prevention. Better techniques and technologies are required to identify unambiguously environmental contributors to health and disease. Standards for large-scale phenotyping (such as in EMRs) would go a long way toward achieving this goal, as would novel approaches to capture data from very large numbers of people.

(Note: NIH has recently launched a CF program called the HMO Collaboratory[13] that will feature a data coordinating center across the HMO research network. In addition, the National Human Genome Research Institute (NHGRI) is already partially sponsoring similar research through the Electronic Medical Records and Genomics[14] (eMERGE) consortium; however, the funding is for specific directed genotype-phenotype studies at only a select few medical centers.)

One example is the large-scale "phenotyping" paradigm[15] currently pursued by the company 23andme, which employs web-based survey instruments to collect data from volunteers. Another is supporting clinical trials that assess the application of pharmacogenomic knowledge on relevant clinical outcomes. In all cases, appropriate attention to ELSI issues is paramount.

Systematic annotation and analyses of GWAS data sets, including non-coding elements, could help narrow a study's "hit" list of candidate loci. Collection, standardization, and integration of data would inform the study of pathways common to various diseases as well as help to tease apart the effects of genes and the environment on health and disease. Validation of GWAS/clinical findings is necessary and could be facilitated by rapid testing abilities. Improved metrics to gather genotype-phenotype data in randomized clinical trials would add significant value to these well-designed studies that inform human health research.

Open access is a major driver of innovation. Despite rules in place to promote data-sharing, this is still a roadblock in GWAS and related studies. Stricter enforcement of data-sharing rules, in concert with cultural change within the scientific community, is needed for progress in this realm.

II. Microbiome and Global Health

[13] http://commonfund.nih.gov/hmocollaboratory/
[14] https://www.mc.vanderbilt.edu/victr/dcc/projects/acc/index.php/Main_Page
[15] https://www.23andme.com/research/

In this session, the discussion leader presented and discussed one paper, summarized briefly below.

1. Bifidobacteria can protect from enteropathogenic infection through production of acetate.[16]

Discussion leader Dr. Howard Hang, The Rockefeller University, provided the following summary:

This paper describes the specific genes from the commensal microorganism *Bifidobacteria* that are important for acetate production, which confers protection from infection by bacterial pathogens such as *E. coli* O157:H7. The study highlights the complexity of commensal-pathogen interactions with host immunity. Dissecting these mechanisms will provide important insights into how host, microbiome, and diet can influence infection. This paper highlights the need for more basic genetic and microbiology studies in commensal bacteria and simpler model systems for dissecting individual commensal-pathogen host interactions.

Session Discussion

The HMP is well under way toward mapping the flora that occupy several sites within the human body. However, while the HMP has provided important funding for large-scale sequencing projects to determine the composition of bacteria in various tissues, future funding on mechanistic studies will be crucial for understanding how commensal bacteria can actually confer protection against disease. The microbiome is a crucial component of gene-environment interactions and their impact on human susceptibility to disease: A role for the microbiome has been suggested in autoimmune, infectious, and metabolic diseases.

Many unanswered questions remain, such as those that explore more deeply microbial-microbial and host-microbial interactions, and the definition of microbial must be broad, to include bacteria, viruses, and parasites. Combinatorial interactions are poorly understood and difficult to study. For these and other reasons, many question the relevance and appropriateness of current model systems such as mice for studying the human gut mucosa. Simpler model systems such as insects, worms, and zebrafish harbor distinct advantages (such as tractable genetics and short breeding times) for investigating problems related to human health. More study of the human innate immune response is also needed to better understand and manage a range of health conditions.

It is becoming clear that diversity of the human microbiome is immense: Better, more efficient tools are needed to create reference maps and standards toward the development of facile “comparative microbiomics.” Although HMP efforts are under way toward accomplishing this goal, more emphasis on functionality and phenotype is the next step. Antibiotic resistance is a key issue in public health, and results from the HMP might be leveraged to address it. Chemical screening of small libraries may avoid killing commensal microbes with antibiotics; alternatively, commensal microbes could be engineered to enable resistance, allowing them to persist and promote health.

[16] Fukuda S, Toh H, Hase K, Oshima K, et al. Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature*. 2011;469:543-7.

The highlighted paper in this session, along with research in this field in general, brings to mind the use of probiotics to promote health. A deeper understanding of host-microbial interactions may lead to such approaches. Indeed, it is important to note that probiotics have been used in this sense for decades, but they have not been evaluated carefully. Whether microbiome findings can be quickly translated into clinical care — through, for example, dietary supplementation — is an open question whose answer hinges on many variables and is not necessarily in the purview of the CF. Collectively, meeting the challenges ahead in this field requires intensely multidisciplinary approaches and interactions. To date, in this field and others similar to it like oral health, mixing tools, techniques, and expertise has yielded the most progress, but more is needed.

III. Group Effects (combinations of infectious and environmental agents, drugs)

In this session, discussion leaders presented and discussed three papers, summarized briefly below.

1. Leishmania RNA virus controls the severity of mucocutaneous leishmaniasis.[17]

2. Discovery of swine as a host for the Reston ebolavirus.[18]

Discussion leader Dr. Erica Ollmann Saphire, The Scripps Research Institute, provided the following summary:

In traditional microbiology dogma, one pathogen causes one disease. However, in the natural environment in which hosts and microbes coexist and collide, there are a multitude of microbes that may co-infect a host — with dampening, additive, or multiplicative effects. In the first paper, researchers found that infection by *Leishmania* parasites, that themselves are infected by a tiny virus, cause exacerbated disease in a mammalian host. Specifically, it is the double-stranded RNA signature of the virus inside the protozoan parasite that confounds the mammalian immune system and allows the parasite to become persistently infected. Detection and analysis of co-infecting microbes would enhance our ability to predict, analyze and defend against disease. A second, surprising example, in which *Reston ebolavirus* was detected in swine co-infected with a typical swine respiratory virus raises the possibility of amplification of a pathogen of human concern by co-infection of nonhuman hosts with nonhuman pathogens. The utility of these studies for the microbial field in general is that they underscore how important it is to consider the possibility that more than one pathogen may contribute to a disease, as well as the likelihood of differential pathogenesis arising from the combination of co-infecting microorganisms.

3. The ecoresponsive genome of Daphnia pulex.[19]

---

[17] Ives A, Ronet C, Prevel F, Ruzzante G, et al. Leishmania RNA virus controls the severity of mucocutaneous leishmaniasis. *Science*. 2011;331:775-8.

[18] Barrette RW, Metwally SA, Rowland JM, Xu L, et al. Discovery of swine as a host for the Reston ebolavirus. *Science*. 2009;325:204-6.

[19] Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, et al. The ecoresponsive genome of Daphnia pulex. *Science*. 2011;331:555-61.

Discussion leader Dr. Julia Gohlke, University of Alabama at Birmingham, provided the following summary:

The draft genome of *Daphnia pulex* (the water flea) introduces an important biomedical model organism that has the potential to develop into a particularly useful whole-animal high-throughput screening base due to its size, transparency, short generation time, large brood size, ease of manipulation, and unique ecology. This paper has the potential for high impact in health research because: i) it describes specific genetic architectures such as co-expansion of gene families via gene duplication, offering a mechanism of rapid evolution explaining the high phenotypic variation across divergent environmental conditions, ii) it analyzes gene expression under varying environmental stressors and shows divergent expression patterns of paralogs after challenge with environmental stressors, and iii) it demonstrates the large number of lineage-specific genes in *Daphnia* responsive to different environmental conditions, including members of the sphingolipid biosynthesis and sex hormone pathways. Compared with other biomedical model organisms, *Daphnia* may be a particularly relevant model for studying environmentally mediated diseases, and drug and chemical screening.

Session Discussion

Confronting the research problem of co-infection and multivariate interactions between host, pathogen, and commensal microorganisms is but one example of the need for systems biology approaches in biomedicine. In toxicology, for example, a similar problem presents as the need to understand the biology and pathology of mixtures. Most studies to date, for explainable reasons mostly due to technology gaps, have focused on the analysis of single compounds, usually from animal studies at very high doses. Although some progress has been made toward studying interactions between drugs and environmental pollutants, still, most of these studies have been limited to a few variables, not many.

In addition to furthering systems biology approaches, new model systems are needed to evaluate group effects — in particular, physiological impact in multiple dimensions (time and space). A key question is, "What constitutes a model system: its criteria, threshold, and/or ability to mimic *in vivo* effects?" Better models are needed for co-infection; one possibility is synthetic biology approaches that build simple organisms with defined properties to evaluate in combination. New model systems are also needed to understand the role of environment, both micro- and macro-, and in particular, those that represent whole organisms in their natural ecology. "Environment" should be considered in its broadest sense.

Model systems should also be chosen deliberately for drug discovery as opposed to basic biology learning, because of the separate goals involved. Many see the NIH as an important resource for providing access and information about the range of model systems available to probe health and disease. One idea is for the NIH to provide a cataloging function of model organisms across disciplines and diseases.

A key bottleneck in the area of "group effects" is the difficulty in ascertaining cause and effect; that is, in identifying biological mechanisms. Better, quantifiable assays would help, as would improved analytical tools that take into account combinatorics and that can manage and dissect multiple variables. 'Omics technologies are needed for large-scale measurement of mixtures: for example, the ability to perform

multiplex RNA interference and single-cell proteomics/metabolomics — especially in "actual" *in situ* environments. The *Daphnia* model (presented in the context of one of the discussion papers) may offer some advantages, although this remains to be tested. Although it is traditionally used as an ecological model system, *Daphnia* has distinct advantages for biomedicine: it is clonal, transparent, has a short generation time (48 hrs), short time to reproductive age (96 hrs) and life span (50 days); and tractable environmental genomics.

Note: Use of the Daphnia *research model is being pursued through activities of the* Daphnia *Genomics Consortium*[20] *and other organizations. Ongoing efforts are focusing on building a* Daphnia *genomics toolbox to be made publically available, facilitating collaborative cross-disciplinary investigations, developing bioinformatic strategies for organizing the rapidly growing genome database, and exploring emerging technologies to improve high throughput analyses of molecular and ecological samples. The genome of* Daphnia pulex *has been sequenced and linkage map exists to facilitate genotype-phenotype studies.*

As with the previous session (microbiome), meeting the challenges ahead in this field requires intensely multidisciplinary approaches and interactions. In a similar vein, borrowing tools and approaches from disparate fields (e.g., astronomy) may be useful for shedding new light on old problems.

IV. Artificial Organs as Tools for Translation

In this session, discussion leaders presented and discussed three papers, summarized briefly below.

1. Reconstituting organ-level lung functions on a chip.[21]

Discussion leader Dr. Andrea Armani, University of Southern California, provided the following summary:

This paper presents a method to fabricate a microsystem that replicates lung function. This specific microsystem will enable researchers to perform experiments, such as studying the uptake of nanoparticles in the lung, without using living systems (animals). There are several potential impacts of this research. On the research or clinical side, these include: i) the immediate impact of the present system; and ii) the possibility of expanding into other organ systems (stomach, intestine, etc). On the ethical side, there is currently a significant debate about the side effects of nanoparticles and nanotechnology in general. Such microsystems would aid researchers in determining if nanoparticles are able to move across different biological barriers (blood/brain, lung, etc). This resulting information could help shape future policy decisions.

2. Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro.[22]

[20] https://daphnia.cgb.indiana.edu/Home

[21] Huh D, Matthews BD, Mammoto A, Montoya-Zavala M et al. Reconstituting organ-level lung functions on a chip. *Science*. 2010;328:1662-8.

[22] Spence JR, Mayhew CN, Rankin SA, Kuhar MF. Directed differentiation of human pluripotent stem cells into ntestinal tissue in vitro. *Nature*. 2011;470:105-9.

Discussion leader Dr. Ophir Klein, University of California, San Francisco, provided the following summary:

A central goal of regenerative medicine is the ability to guide pluripotent stem cells to form functional three-dimensional organs, but such strategies seemed the stuff of science fiction only a few years ago. This paper reports the generation of three-dimensional, intestinal 'organoids' from pluripotent (ES or iPS) cells by directed differentiation with specific factors through a series of defined intermediates. This paper, like several others published over the past year, builds beautifully on data acquired from embryonic studies about the signaling pathways that regulate successive fate decisions during intestinal development. The authors use this knowledge to coax the cells to progressively acquire more differentiated states, but in contrast to previous studies, they are able to obtain three-dimensional structures whose architecture closely resembles that of native tissue. Replacement organs represent promising treatments for myriad diseases, including diabetes, neurodegenerative disorders, heart or kidney failure, as well as for trauma and congenital anomalies. Thus, the production of bioengineered tissues or organs will have major impacts on most fields of medicine, particularly as the population ages and organ failure increases as a cause of morbidity.

3. Patient-specific induced pluripotent stem-cell models for long-QT syndrome.[23]

Discussion leader Dr. Joseph Wu, Stanford University School of Medicine, provided the following summary:

Long-QT syndromes are heritable diseases associated with prolongation of the QT interval on an electrocardiogram and a high risk of sudden cardiac death due to ventricular tachyarrhythmia. In long-QT syndrome type 1, mutations occur in the KCNQ1 gene, which encodes the repolarizing potassium channel mediating the delayed rectifier I(Ks) current. This study screened a family affected by long-QT syndrome type 1 and identified an autosomal dominant missense mutation in the KCNQ1 gene. The researchers obtained dermal fibroblasts from two family members and two healthy controls and infected them with retroviral vectors encoding the human transcription factors OCT3/4, SOX2, KLF4, and c-MYC to generate pluripotent stem cells, which were then directed to differentiate into cardiac myocytes. The resulting iPS cells maintained the disease genotype of long-QT syndrome type 1 and generated functional myocytes. Individual cells showed a "ventricular," "atrial," or "nodal" phenotype, as evidenced by the expression of cell-type–specific markers and as seen in recordings of the action potentials in single cells. The patient-derived cells recapitulated the electrophysiological features of a long-QT syndrome. This paper has the potential for high impact because it uses iPS cells as a drug screening platform.

Session Discussion

As with the previous sessions, there is strong agreement that investing in interdisciplinary conceptual integration will stimulate innovation. Tissue engineering is an attractive area for many reasons, not the

[23] Moretti A, Bellin M, Welling A, Jung CB, et al. Patient-specific induced pluripotent stem-cell models for long-QT syndrome. *N Engl J Med*. 2010;363:1397-409.

least of which is its near-term clinical applicability for accelerating drug development across several systems and disorders. Combining this type of technology with advances in pharmacogenomics may be especially productive. The CF may be a useful vehicle to build bridges not only for the individuals with big ideas, but also to facilitate better partnerships with enterprising businesses to co-fund such advances, even if a few such endeavors already exist. As has been stated previously, the assembly of biologists and quantitative scientists is made ever more difficult with persistent roadblocks in peer review.

Note: many NIH efforts have been undertaken, and/or are under way, to play "matchmaker" between biologists and quantitative scientists, with varying levels of success. The pipeline is a significant issue, as is communication: getting the word out about what the NIH is already doing.

There is considerable enthusiasm for simulating the development of major organ systems using human stem cells (iPS cells in particular) to identify disease mechanisms, improve outcomes for complex disorders, as well as help to design preventive strategies. There is even more enthusiasm for combining the use of stem cells with bioengineering approaches that offer a level of predictability and quantitation not common to nearly all biological systems (especially those more complex than individual cells). It should be noted that while three-dimensionality is extremely important in tissue engineering, the term has distinct meanings (e.g., for screening, for investigation, for implantation in humans). Whether some of the advances in this field represent natural progress or truly innovative thinking remains unclear, but they do represent important science that is of great interest to the NIH mission and public health.

Artificial organs, organoids, and similar biomimics are likely to find wide use in biomedicine. The ability to create a specific disease model in the lab is very powerful, but one issue that still requires attention is developmental trajectory: tracking drug/pathogen effects, tissue development/differentiation/aging over time. This is not readily doable with current model systems. Another important step, which may be facilitated by a CF mechanism, would be to derive and post common pathways across diseases — an endeavor not unique to any IC but still immensely useful to all disease investigations.

V. Proteomics and Therapeutics Development

In this session, discussion leaders presented and discussed two papers, summarized briefly below.

1. Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes.[24]

Discussion leader Dr. Ileana Cristea, Princeton University, provided the following summary:

Understanding protein interactions and activities in a spatial and temporal context provides the opportunity to design therapeutic interventions targeted to a specific protein function. Yet the majority of proteins are expressed in most tissues and cell types, meaning that it is not just the on/off expression of a protein that determines a phenotype or function. Numerous proteins are part of different protein complexes with diverse functions, and the localization, interactions, posttranslational modifications and

[24] Bantscheff M, Hopf C, Savitski MM, Dittmann A, et al. Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nat Biotechnol*. 2011;29:255-65.

relative expression levels contribute to both protein and protein complex function. Recent reports have emphasized the importance of the epigenetic landscape of a cell in the development and susceptibility to disease. Chromatin remodeling enzymes critically contribute to the establishment of this epigenetic landscape through finely-tuned activities of multi-protein complexes. Histone deacetylases (HDAC) are such enzymes that are involved in both gene transcription and its epigenetic regulation. HDAC inhibitors (HDACi) are used as therapeutic reagents, including compounds in clinical use for certain cancer treatments. HDACi have also stirred interest in studies and treatment of infectious diseases, as their use was shown to trigger reactivation of certain viruses from latency (e.g., HSV1, HCMV, HIV-1), latency constituting one of the main limiting factors of therapeutic success. This research identified drug-specific targets and associated HDAC-containing complexes with distinct functions using a multidisciplinary approach integrating: chemoproteomics (competition for binding to HDACi), isolation of protein complexes, quantitative mass spectrometry (isobaric tags and LC MS/MS analyses) and cellular fractionation. The researchers demonstrated that drugs have different affinities toward different protein complexes containing the same catalytic subunit. This paper shows that it is at the protein complex level that an enzyme exerts its specific and unique function. It indicates that a protein complex may be more specific and effective target for therapeutic intervention than a single protein, and can help decrease the number of secondary effects following drug treatment. It also demonstrates the power of multidisciplinary approaches for obtaining a more complete view of a protein function.

2. Protein disulphide isomerase protects against protein aggregation and is S-nitrosylated in amyotrophic lateral sclerosis.[25]

Discussion leader Dr. Desiré Tshala-Katumbay, Oregon Health and Science University/University of Kinshasa, DR Congo, provided the following summary:

Understanding mechanisms of neurodegenerative diseases has remained a challenge in modern biomedicine despite current advances in investigative neuroscience. In this paper, the researchers used a discrete chemical tool (siRNA), in combination with functional proteomics, to establish the foundation of novel research lines that can help develop therapies for neurodegenerative diseases — notably those associated with protein misfolding such Alzheimer disease, Parkinson disease, and amyotrophic lateral sclerosis. This research: i) identified protein disulfide isomerase, a member of the multigenerational thioredoxin superfamily, as a major player in neurodegenerative mechanisms associated with protein misfolding; ii) demonstrated the neuroprotective properties of protein disulfide isomerase; iii) elucidated the structure-activity requirements for other molecules to (1) display protein disulfide isomerase-like neuroprotective properties; and iv) established the foundation for thiol-redox regulation of protein folding to be tested as a therapeutic strategy for neurodegenerative diseases associated with protein misfolding.

Session Discussion

[25] Walker AK, Farg MA, Bye CR, McLean CA, et al. Protein disulphide isomerase protects against protein aggregation and is S-nitrosylated in amyotrophic lateral sclerosis. *Brain*. 2010;133:105-16.

Mapping the "complexome" (e.g., protein-protein, protein-nucleic acid, protein-lipid, protein-carbohydrate) would have high value across ICs and diseases, and thus is a project relevant for CF investment. It may also play a role in unveiling the "dark matter" in drug development. Taking a "drug-centric" approach may enable a systematic way to uncover new uses for existing drugs. Such a "cellular control system" could enable rational drug design in a much more systematic way than is currently possible. Another potential application for complexome analyses would be the opportunity to study polypharmacology — perhaps in combination with organoids/artificial organs. Currently, there is a dearth of options for studying the *in vivo* effects of drugs used in combination. Engaging industry in any such efforts will be critical for maximizing the potential of this type of research.

Note: The NIH is in discussion with industry about the use of "discarded" compounds, and it is hoped that many of these activities will take place in the new NCATS, which is slated to open in October, 2011.

The concept of specifically targeting a transient protein complex moves beyond the paradigm of targeting a single protein's catalytic site that currently dominates the field of drug discovery. One issue to consider, however, is protein/protein complex heterogeneity — as it occurs in disease states or even in normal functioning. Many studies have shown that proteins are highly dynamic entities, and this issue must be contended with in some fashion.

Once again, the issue of team science, and of bringing new fields and new ways of thinking together, is an important concept in proteomics and novel therapeutic development.

VI. Single Cell Analysis

In this session, discussion leaders presented and discussed two papers, summarized briefly below.

1. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations.[26]

Discussion leader Dr. Brad Bernstein, Massachusetts General Hospital, provided the following summary:

This paper describes a subpopulation of lung cancer cells that are resistant to drug therapy. The researchers present evidence that the resistant sub-population relies on distinct signaling pathways and a distinct chromatin state. Moreover, they demonstrate that the changes are reversible, suggesting an epigenetic mechanism. This paper highlights a central problem in oncology — namely, that tumors acquire drug resistance through genetic (or in this case epigenetic) alterations of a subset of population (which could be rare cells).

2. Bacterial charity work leads to population-wide resistance.[27]

Discussion leader Dr. Angelique Whitehurst, University of North Carolina, Chapel Hill, provided the following summary:

[26] Sharma SV, Lee DY, Li B, Quinlan MP, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*. 2010;141:69-80.
[27] Lee HH, Molla MN, Cantor CR, Collins JJ. Bacterial charity work leads to population-wide resistance. *Nature*. 2010;467:82-5.

The seminal finding of this study is that a minority of cells can confer antibiotic resistance to an entire bacterial population through a paracrine signaling mechanism that involves the production of indole to activate drug transporters. Extrapolating to a broader context, this work demonstrates elegantly how observations at a single-cell level can reveal the underlying basis for population behavior that in turn presents opportunities for identifying therapeutic intervention strategies that would otherwise not be detected. Thus, this study clearly highlights how the simultaneous analysis of systems at multiples levels of resolution can reveal new discovery space in our understanding of complex human disease.

Session Discussion

Note: The NIH Common Fund actually is already considering an initiative in single-cell analysis.

Single-cell analysis is a research endeavor that bridges most fields and interests in biomedicine, and its value extends beyond fundamental research or technology development. One main reason is the need to come to terms with the fact that most studies of cells and their behavior have been performed on populations of cells that we now know are, in most cases, highly heterogeneous. Thus, many standing conclusions in biomedicine have been based upon averages of ensembles of cells.

Moreover, individual cells in a population acquire “extra” behavior; knowing how to identify and predict these changes is broadly valuable for understanding the cellular and organismal systems that constitute health and disease. Recent technological advances suggest that the time is right for expanding investment in the ability to monitor individual cell contents and behavior, both alone and within populations and sub-populations. The ability to analyze single cells and their cognate populations concomitantly would also be valuable in helping to clarify the conditions and perturbations that determine cellular “decision-making.” Sophisticated models that can identify and distinguish between sub-groups within a cellular population may shed light on important biological problems such as antibiotic resistance, tumor development and spread, infection, and many others. One paper presented in this session that shows reversible drug resistance is an important step in this direction.

The ultimate goal of this field of research is the ability to monitor (and ultimately, manipulate), *in situ*, all of the parameters of single-cell behavior within populations and subpopulations. Doing so would remove potentially artifacts caused by over-expression or other artificial conditions of cell culture and maintenance, and the impact would likely be felt by virtually all ICs, diseases, and disciplines.

As with previously discussed topics, meeting the challenges ahead in this field requires intensely multidisciplinary approaches and interactions — in particular, true teamwork between biologists and quantitative scientists. However, there appears to be enormous barrier to entry into this field for engineers and researchers who aren’t clinicians or biologists, as their labs are not equipped to perform experiments with “real” samples. Moreover, initiating collaborations across fields (engineering and medicine, for example) is much harder than across different fields of engineering or different fields of medicine, because of simple language barriers. Therefore, it is necessary to engage technologists in these discussions.

The computational challenges in detecting different behaviors in sub-populations of cells remain significant. There appear to be similarities of single-cell and population analyses with social network analyses (e.g. covert networks). A vast literature on organizational theory may be relevant to pursue and apply to single-cell analysis.

VII. NIH Award Strategies

In this session, discussion leaders presented and discussed two papers, summarized briefly below.

1. Optogenetics.[28]

Discussion leader Dr. Teri Odom, Northwestern University, provided the following summary:

As a field, optogenetics has potential for high impact because in its stripped-down form, it enables the study of well-defined events at specific times in whole organisms. After gene modification to insert a light-responsive receptor, controlled cellular signaling can be interpreted within the context of its native tissue environment and even be correlated with behavior. There are other possibilities to understand brain-related diseases as they are associated with complex tissue function and targeted neuron populations. Besides the science and potential of this new field, optogenetics is exciting because it requires expertise from and integration of teams across several different areas (optics, genetics, neuroscience, psychology) to be fully realized.

2. Glass-like dynamics of collective cell migration.[29]

Discussion leader Dr. Christine Payne, Georgia Institute of Technology, provided the following summary:

Wound healing, embryonic development, organ regeneration, and many other aspects of human health depend on the collective motion of cells. To initiate, control, and correct these events requires a fundamental understanding of cell migration. This paper provide a significant step toward understanding the collective motion of cells by recording time-lapse images of a monolayer of cells on a tissue-like substrate. The researchers found that cell migration slows as cell density increases, but that the fastest moving cells are clustered in increasingly larger groups. This behavior is similar to the dynamic heterogeneity observed for colloidal or particulate systems as they approach the glass transition. Additional glass transition-like behaviors were observed in the monolayer of cells, strengthening the argument that migrating cells at high cell densities behave similarly to non-biological systems near the glass transition.

This paper, which connects an important biological phenomenon to a general physical principle, has the potential for high impact due to its fundamental nature, allowing it to be applied to many different health-related questions. While translational research has important applications for human health, fundamental research such as that described in this work has the potential for much greater impact because it is general. The strength of such research is that a fundamental and well-studied physical

[28] Deisseroth K. Optogenetics. *Nat Methods*. 2011;8:26-9.
[29] Angelini TE, Hannezo E, Trepat X, Marquez M, et al. Glass-like dynamics of collective cell migration. *Proc Natl Acad Sci U S A*. 2011;108:4714-9.

phenomenon, the behavior of glassy systems, can now be applied to a range of important biomedical questions. In this sense an analogy can be made with synthetic organic chemistry and drug development. Although a specific drug might not make it to market, the synthetic methodologies used to develop that drug can be applied to a range of other drug targets.

Session Discussion

Both papers presented in this session speak to the need for bringing together innovative people from distinct fields of science. Many view the most successful cross-disciplinary collaborations as being self-assembled and not too large. A key bottleneck appears to be finding collaborators, since investigators within different disciplines attend different sets of meetings, apply for different funding, read different literature, and speak different "languages." Suggestions for the NIH include providing a formal mechanism to encourage substantial research collaborations between scientists carrying out fundamental research and those working on specific biomedical problems. Examples include:

- Creating biomedical challenges that request a description of outstanding research questions from biomedical researchers and then solicit proposals for new approaches or solutions from researchers not typically funded through the NIH. Funding would go to collaborations between the pairs (or small groups).
- Sabbatical funding for non-biomedical researchers to work in a biomedical lab. Non-NIH scientists offer the potential for new instrumentation (that can be difficult to fund outside the NIH) and approaches to biomedical questions. Embedding them in biomedical research lab would expose them to the questions and techniques in the biomedical field and foster substantial collaborations.
- DARPA[30]-like approach in which the NIH identifies an important biomedical problem, then populates and manages a diverse team to solve it in a defined time-frame.
- Funding explicit training for quantitative scientists aiming to enter biomedical research.

  (Note: the NIH funds K25 Mentored Quantitative Research Development Awards[31] for this purpose)

- Funding pilot grants of collaborations that proceed, if successful, to later-stage, or "mezzanine" awards.
- "Boot camps," in which potential collaborators from different fields are pre-selected, invited to a workshop, offered a demonstration project/data set, and asked to write a grant proposal as a deliverable.

Although the NIH has designed and implemented several strategies that aim to bring together bright minds from different disciplines, better communication with the scientific community, along with improved awareness of the NIH, may help to connect extant programs with applicants. One suggestion

[30] http://www.darpa.mil/default.aspx
[31] http://www.nigms.nih.gov/Training/CareerDev/MentoredQuantResQA.htm

is “pre-events,” in which the NIH hosts a meeting/workshop/online space that publicizes a program in advance, attracting applicants and helping to connect potential collaborators.

Whether or not the NIH should fund foreign investigators as trainees also remains an ongoing issue of interest and concern within the scientific community. By law, non-U.S. citizens are ineligible for most NIH-funded post-doctoral training funds. However, there are exceptions, one being training associated with the NIH CF Interdisciplinary Research project,[32] which is available to all NIH ICs.

Teaming with stakeholders is also important. Suggestions include: i) enabling the NIH to act as venture capital to fund intermediate-level applied science unattractive to venture capital and industry, ii) authenticating institutional seed funding with the NIH “stamp of approval,” and iii) enabling partnership with industry early in the research stage to facilitate later translation.

VIII. Other

In this session, discussion leaders presented and discussed six papers, summarized briefly below, and discussed miscellaneous topics.

1. Protease cleavage sites in HIV-1 gp120 recognized by antigen processing enzymes are conserved and located at receptor binding sites.[33]

Summary: This study showed that antigen processing protease cleavage sites in the envelope protein of HIV-1 (gp120) cluster in proximity to conformational antibody targeted epitopes. It is the first structural evidence that speaks to the phenomenon of “linked recognition” in immunology, whereby an antibody response to a conformational epitope in a target molecule can only be mounted by a B-cell if a T-cell reacts with a different linear (cryptic) epitope from that same target molecule presented by an MHC molecule on an antigen-presenting cell. This understanding may underlie the current bottleneck in all vaccine development — rational elicitation of specific antibodies by immunogens (broadly neutralizing epitopes are known to be present on gp120 and HIV positive patients exhibit antibodies to them but exact structural mimics of these epitopes have always failed to elicit the antibodies on demand), in all allergy/autoimmune diseases — why autoantibodies are elicited in some people but not others for the same autoantigen, in cancer — why antibodies do not form to the aberrant surfaces of cancer cells, in various infectious diseases and probably in many diseases with inflammation as one of their components.

2. Leptin therapy in insulin-deficient type I diabetes.[34]

Summary: Leptin is a hormone that regulates energy intake and expenditure among a myriad of other functions. It was discovered in 1994, but it has taken close to 20 years to test for its effect on sugar control. This study tests recombinant leptin injections in a mouse model of diabetes. The therapeutic

[32] http://commonfund.nih.gov/interdisciplinary/consortia/index.aspx

[33] Yu B, Fonseca DP, O'Rourke SM, Berman PW. Protease cleavage sites in HIV-1 gp120 recognized by antigen processing enzymes are conserved and located at receptor binding sites. *J Virol*. 2010;84:1513-26.

[34] Wang MY, Chen L, Clark GO, Lee Y, et al. Leptin therapy in insulin-deficient type I diabetes. *Proc Natl Acad Sci U S A*. 2010;107:4813-9.

effects were superior to insulin in terms of lipid and cholesterol control, while glycemic control was similar (if not slightly better) for leptin compared to insulin. This suggests that leptin would work better than insulin. The study also makes the remarkable assertion that mice without functional pancreatic beta cells can control glycemia. Hitherto, glucose levels could only be controlled via manipulating insulin and glucagon (or dietary approaches).

3. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis.[35]

Summary: This paper describes the identification of blood transcriptional signatures that allow classification of active versus latent tuberculosis, and the ability to distinguish these disease states from other non-tubercular disease (infectious or inflammatory) in humans. By measuring a median "molecular distance to health" metric that is a composite of the number of transcripts in a profile that differ from a healthy control, the researchers were able to distinguish patients with advanced disease, minimal or no disease, and track response to treatment. The signatures reflect a combination of alterations in cell population and changes in expression within particular cell types. Further, they are able to implicate an interferon-inducible neutrophil response and identify a role for type I IFN-ab signaling in tuberculosis pathogenesis.

4. Selective cell death mediated by small conditional RNAs.[36]

Summary: This paper demonstrates an RNA-based mechanism for the tumor-cell specific activation of an innate immune response leading to cell death. The main impact of this paper comes from the demonstration that an engineered, multi-step reaction pathway that does not rely on existing biological components can work reliably in a complex cellular environment and can even interface with an existing biological pathway. The reaction mechanism — the controlled formation of long double-stranded RNA polymers from short single-stranded hairpin monomers —was completely rationally designed and was initially developed and tested in a cell-free setting.

5. A common allele in the oxytocin receptor gene (OXTR) impacts prosocial temperament and human hypothalamic-limbic structure and function.[37]

Summary: This paper represents a multi-modal approach combining genetic, phenotypic, and neuroimaging data to study oxytocin, an evolutionarily conserved neuropeptide with a wide variety of physiological functions in the brain and the body. Specifically, a functional oxtyocin receptor polymorphism linked to social functioning predicted individual differences in hypothalamic limbic brain circuits and pro-social behavior in healthy humans. Furthermore, structural changes in the hypothalamus in males with this polymorphism predicted impaired pro-social temperament suggesting

---

[35] Berry MP, Graham CM, McNab FW, Xu Z, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature*. 2010;466:973-7.

[36] Venkataraman S, Dirks RM, Ueda CT, Pierce NA. Selective cell death mediated by small conditional RNAs. *Proc Natl Acad Sci U S A*. 2010;107:16777-82.

[37] Tost H, Kolachana B, Hakimi S, Lemaitre H, et al. A common allele in the oxytocin receptor gene (OXTR) impacts prosocial temperament and human hypothalamic-limbic structure and function. *Proc Natl Acad Sci U S A.* 2010;107:13936-41.

a sex-dependent impact of this genetic variant. Collectively, the findings suggest a neural mechanism linking genetic risk in oxytocin functioning and individual differences in emotional reactivity and pro-social temperament. The paper also nicely highlights caveats in the field, namely the need to have more coordinated studies testing similar concepts and constructs using similar methods across a diversity of conditions, both healthy and pathological states, in order for humans to make progress in the etiology, treatment and eventual prevention of human disease. This type of study also underscores the idea that biology and experience, especially involving social interactions, are bidirectionally influenced, raising the possibility that empathic therapeutic interventions can have as great if not greater impact on gene expression as biological/pharmacological interventions or at least provide synergistic enhancement.

6. Reversing pathological neural activity using targeted plasticity.[38]

Summary: This study makes remarkable use of neuroplasticity to alter neural responses and restore behavioral function in an animal model of tinnitus. Behavioral therapies have long made use of knowledge that the brain is plastic, but these therapies have treated the brain as a "black box." Modern neuroscience has expanded our understanding of the function of neural circuits, as well as of mechanisms by which they can be strengthened and weakened. Advances in imaging have improved the ability to measure these changes longitudinally in humans. Combinations of deep brain stimulation, vagal stimulation, or short-term pharmacology with focal behavioral exercise might be used productively in the future to rationally sculpt circuits to treat psychiatric and neurological diseases. Identification of neural markers of resilience could also guide preventative behavioral or mild pharmaco-behavioral therapies to sculpt resilient brains in at-risk populations.

Session Discussion

The wide range of topics covered in this last session invited several comments not necessarily following a common theme. These comments are listed below with relevant contextual discussion, if applicable:

1. Does the NIH have the ability to act quickly on research findings that show exceptional promise for improving public health?

Many view the translational process as long, cumbersome, and marginally effective. Although the NIH plays a direct role in some aspects of the process, many factors are in play and are outside the agency's purview. These include third-party payer policies and other aspects of the U.S. healthcare system, unpredictable industry investment and follow-up, and behavioral change (individual, community, society) that is also unpredictable and difficult to influence. Nonetheless, the NIH does monitor ongoing research and uses various approaches to try to move results more quickly to the clinic. For example, the NIH has ongoing discussions and collaborations with sister health agencies such as the Center for Medicare and Medicaid Services, the Agency for Healthcare and Research Quality, the Centers for Disease Control and Prevention, and the Food and Drug Administration. The issue is extremely

[38] Engineer ND, Riley JR, Seale JD, Vrana WA, et al. Reversing pathological neural activity using targeted plasticity. *Nature*. 2011;470:101-4.

multifaceted and complex, but of clear importance and urgency to both the NIH and its stakeholders, including Congress and the public.

2. The oxytocin paper presented in this session prompted discussion of the importance of enhancing understanding of the brain and behavior.

There is an inevitable "catch-22" for humans to study their own thinking and behavior, and thus these investigations are extraordinarily complex. Model systems that analyze conserved pathways and molecules across multiple levels, disease states, and lifespan (and that measure intermediate phenotypes) may begin to shed light on "being human." The PROMIS behavioral instruments[39] that have resulted from CF-supported research are excellent tools for investigating mind-body interactions. The NIH could encourage more interdisciplinary use of the PROMIS instruments to understand better understand a wider range of diseases and processes.

3. Innate immunity is a topic of broad interest and one that is likely to underlie many aspects of health and disease. Its study is thus relevant for CF investment, through large-scale studies as well as cell- and tissue-specific investigations.

4. Personalized medicine has become an overused term, but the concept remains an extremely important application of the large-scale collection of genomic, proteomic, and other 'omic data that is ongoing.

The ability to apply molecular diagnoses that stratify patients for disease severity, treatment, or outcome, as is exemplified by the presented paper on tuberculosis, is an exciting prospect that requires more large-scale study. "Pseudo-phenotypes" may be useful in sub-classification of health conditions, which could lead to improved pharmacotherapy.

5. Is there/what is the role for the CF in prevention research?

This question also addresses a complex issue, since effective prevention is often many steps removed from initial research. However, the NIH works with its partners (foundations and advocacy organizations, industry, communities, policy makers, and the international community) toward implementation of research findings into real settings. In addition, the NIH has long supported comparative effectiveness research studies, which compare existing therapies and often inform policy decisions.

6. Although the NIH has issued a formal data-sharing policy, many in the scientific community view its enforcement as spotty, leading to delays in progress. Stricter enforcement of the broad sharing of reagents, biospecimens, and materials would be a welcome step.

7. The NIH should continue its CF investment in high-risk, high-reward research (e.g., Pioneer, New Innovator award programs). Investing in individual creativity has great value as does studying "mysterious," often non-fundable through ordinary means, scientific questions.

[39] http://www.nihpromis.org/

Day 3: Group Summaries and Recommendations

I. Beyond GWAS

Challenges and Obstacles

Although GWAS have uncovered many genetic loci for a range of conditions and diseases, a major challenge is translating this knowledge into functional insights. One key roadblock is the inability to capture precisely various and diverse environmental measurements. Incomplete, nonstandardized, and shallow collection of phenotype data contributes to the difficulty of using GWAS data to define mechanisms and/or suggest potential interventions. Insufficient sample sizes prohibit the clarification of the role and relevance of complex traits in health and disease. In some cases, valuable opportunities may be missed, as in harnessing genotyping data from randomized clinical trials that have rich phenotypic data. For the massive amounts of data that already exist, practical and effective strategies for integration lag behind. Possible remedies include new algorithms for performing higher-order 'omics studies, a repository of rare knockouts, and more complete sharing of data and biospecimens.

Emerging Opportunities

Further progress in GWAS requires both persistence and innovation. While GWAS execution is routine and fairly well-established (in the "**D**" portion of the graph below), others are in a period of rapid growth (in the "**C**" region of the curve: single trait analysis, eQTLs), and still others require a substantial push to reach their potential (in the "**A**" and "**B**" areas below: functional annotation of genetic variants, annotation of a reference genome (ENCODE), whole-genome analyses in unrelateds/families, large-scale phenotyping, and clinical translation). This last group is likely to be the most ripe for CF investment.

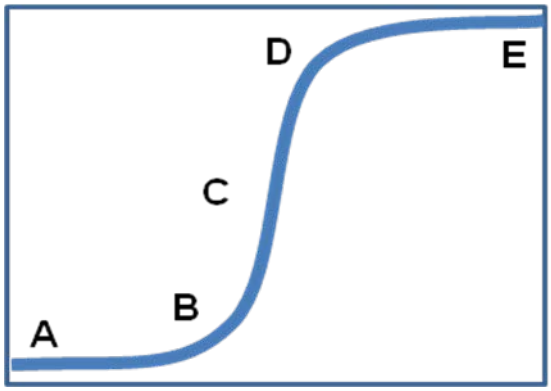

Fig. 1. Scientific progress over time

Potential CF Investments

Three proposed projects (each independent but complementary and potentially synergistic) could overcome some of the current roadblocks in this area.

- Human Phe-Ge Project. This proposed project represents a very large-scale effort to create a "National Cohort" of people (DNA plus phenotypic data) for discovery research in health and disease. The large sample size (1 million people across the United States) would permit sufficient coverage of the human genome, along with a diversity of participants that reflect the U.S. population make-up. Clinical data would be harvested from EMRs (after participants opt in), and the cohort would be followed longitudinally. Web surveys (e.g. 23andwe) could harness the reach and power of social networking to gather data in real settings. Within the larger cohort would be a sub-cohort of approximately 1,000 people, who

would be subjected, upon consent, for deep phenotyping and clinical validation. Data sharing would be free and wide, with appropriate consent in place from volunteer participants.

- Functional Genome Project. This potential project would leverage functional information to find causal variants, employing ENCODE,[40] epigenomics, and functional genomics strategies. Functional annotation of 1,000 individuals over multiple cell types / conditions would record transcription, DNA methylation / histone modifications, and DNA sequencing (phased whole-genome sequencing). The project aims to advance GWAS science by yielding a more granular phenotype that will enable faster translation of genomic findings to clinical applications.
- Multidimensional Analyses for Genomic Studies. To further address the issue of GWAS data integration, this project would strive to provide context for genomic data by accessing environmental measures, incorporating population and family structure, and including epigenetic context. Higher-level interactions could be identified through the capture of functional interactions, pathway analyses, and novel combinatorics approaches. Candidate methodological innovations include such as more flexible analysis methods and study designs, whole-genome sequencing, and computational improvements that speed and expand processing capabilities.

Potential Impact

Moving GWAS beyond its current capability offers faster movement from association to function, which will likely accelerate discovery for multiple traits. Clinical relevance of most GWAS to date is lacking: The proposed projects aim to lead to better clinical decision support, new diagnostics and therapeutics, improved coordination with industry, as well as the realization of meaningful use criteria of the HITECH act.[41]

Session Discussion

The eMERGE Network, recently established through NIH funding, is a national consortium formed to develop, disseminate, and apply approaches to research that combine DNA biorepositories with EMR systems for large-scale, high-throughput genetic research. With additional funding, it could serve as a vehicle for the proposed national cohort study. To realize potential, however, several issues would need to be addressed, including phenotype standardization, careful data reconciliation, and adjunct use of additional data collection streams (such as the web health surveys posted by companies like 23andMe). An extension/component of this proposed project is the retro-analysis of a sub-cohort of older Americans with known diseases, to validate genotyping efforts.

II. Microbiome: Part 2

Challenges and Obstacles

[40] http://www.genome.gov/10005107

[41] The Health Information Technology for Economic and Clinical Health (HITECH) Act, enacted as part of the American Recovery and Reinvestment Act of 2009, was signed into law on February 17, 2009, to promote the adoption and meaningful use of health information technology.

A substantial CF effort to study the human microbiome has yielded many findings and insights. However, moving beyond sequencing to a functional understanding of the microbiome is now the challenge.

Emerging Opportunities

The emerging opportunity to understand the function of the microbiome has been generated by the sequencing data and the early correlation studies funded by the Human Microbiome Project. There are challenges that need to be overcome to exploit these data, but the opportunity exists to fundamentally shift the paradigm that describes the human-microbiome relationship and the impact that the microbiome has on the health of many tissues.

Potential CF Investments

The CF could accelerate the potential transformative impact of the microbiome sequencing data by investing in the following:

- Computational tools that will allow deconvolution of complex data sets
- Functional assays that go beyond sequencing to understand the impact of the microbiome, including metabolomics, glycan-processing readouts, and short-chain fatty acid biosynthetic pathway analysis
- Functional model systems to understand bacteria-host interactions, bacteria-bacteria interactions, and bacteria-virus interactions. These model systems might include iPS-derived organ cultures as well as animal models.
- *In vitro* systems for microbiology and genetics of commensal bacteria
- Functional analysis of small molecules derived from microbes
- Systemic impact of the gut microbiome on other organs
- GWAS-MWAS (Microbiome-Wide Association Studies) research
- Mother-child microbiome effects
- Any NIH-sponsored efforts in this area should be multidisciplinary to realize the fullest impact of potential.

Potential Impact

Systems biology-derived “designer” probiotics may offer an inexpensive, holistic approach to disease prevention and treatment, although much research is necessary to realize this potential. It should be noted that probiotics have been used for decades, but systematic analyses of their efficacy and safety have not been conducted.

Note: Many of the issues described are already under study in the oral health field, which hosts a large body of literature on host, commensal, pathogen interactions and functional consequences on physiology and pathology. One clear example of the “mouth as a laboratory” is its accessibility and the availability of various tissues to compare and study. Proposed studies of the gut microbiome would benefit greatly from visiting the experience and findings of the oral health research arena.

III. Group Effects

Challenges and Obstacles

Exposures are highly variable and dynamic throughout the lifetime of an individual. Needed are systematic, unbiased screens for studying how multiple factors (e.g. microbiological, chemical, lifestyle and dietary exposures) interact to contribute to susceptibility to disease, disease progression, and treatment outcomes. In addition to curating/annotating data obtained using current models, improved testing systems are needed that are equipped to analyze multi-factorial issues.

Emerging Opportunities

Several opportunities exist to address the need for better models and analytic tools. These include the availability of inexpensive exposure screening tools (e.g. virochip, protein adducts) and bioinformatic techniques that can handle large, clinical datasets to track exposures. The development of screening tools, methods, and model systems that are particularly well suited for studying mechanisms of environmental influence also provide opportunities in this area. Point-of-care tools are likely to be especially useful to monitor exposure in global and other low-resource settings. Expanding these toward multiplex capability is another opportunity.

Potential CF Investments

The CF could shift the curve to accelerate progress by expanding the number and quality of tools to systematically measure multiple exposures and by supporting the development of computational tools that will support multifactorial research: viral, bacterial, chemical, and dietary. Data handling for these types of studies is an enormous challenge. A database that catalogs and characterizes model systems that are suitable for studying multifactorial research would also be helpful.

Note: the NIH Genes and Environment initiative — in particular, the Exposure Biology program[42], which is under way, is currently funding projects that encourage the development of innovative technologies to measure environmental exposures, diet, physical activity, psychosocial stress, and addictive substances that contribute to the development of disease. A second round of funding is expected.

Potential Impact

Implementing projects in this area could have significant impact in helping to better clarify the age-old question of the relative influences of "nature and nurture;" yet, it would go further by ultimately explaining how complex mixtures of genetic loci and environmental exposures influence health and disease susceptibility. In time, these insights will point to preventive strategies that help to fulfill the goals of personalized medicine.

IV. Artificial Organs as Tools for Translation

Challenges and Obstacles

[42] http://www.gei.nih.gov/exposurebiology/

Current knowledge of human health and disease is overly dependent on the results of *in vitro* models, which have variable value in replicating human systems. As an improvement over cell culture and some animal models, many researchers are turning to organ-mimics to perform some of these key experiments. Currently, most organ-mimics recapitulate only a fraction of the function and have little of the structure of the primary organ, but it is clear that both structure and function play a critical role in biological behavior. Experiments using organ-mimics also offer the potential to deepen understanding of the systems biology of organs and tissues.

Emerging Opportunities

Emerging capabilities for the development of organ-mimics that maintain both structure and function of the original organ offer the opportunity to overcome limitations of *in vitro* studies that probe human biology. The development of three-dimensional organ-mimics could have short-term impact in drug screening and long-term benefit as replacement organs for the regenerative medicine toolkit.

Potential CF Investments

- Multidisciplinary work and interactions are essential in this field, and thus NIH awards in this area should be developed for both individual investigators and multi-investigator projects. Specific recommendations include:
- Projects that integrate developmental biology, organ-quality assays, and bioengineering should be supported. The frontiers of developmental biology should be explored so that engineering approaches can be better informed.
- The NIH should push for technology transfer mechanisms that accelerate tool- and instrumentation sharing.
- Development of a human iPS cell repository would enable studies of disease pathogenesis as well as those that seek to develop organoids from diverse genotypes.
- The use of iPS cells to develop organoids with diverse genotypes would facilitate the development of more robust drug screening platforms that could distinguish safety/toxicology issues that vary with genotype. For example, these organoids could offer an improvement over the traditional hERG (human Ether-a-go-go Related Gene) test that is used to assess cardiac safety of new drugs.
- The California Institute for Regenerative Medicine hosts a biorepository of iPS cells for the study of neurodegenerative disease; this sort of resource, on a broader scale, would be very useful to the scientific community.
- Establish an NIH-funded “Jackson Lab” for organ-mimics.

Potential Impact

Progress in this area has substantial promise, including a paradigm shift in conducting clinical trials, from costly and slow to cheap and fast. Other benefits of true organ-mimics include: i) pushing the frontier of developmental biology research and ii) accelerating preclinical evaluations (providing a next-generation

"immunoassay"). Long-term benefits include the development and use of artificial organs, which are of obvious relevance for thousands of diseases and conditions.

V. Targeting the Dynamic Complexome

Challenges and Obstacles

The spatial and temporal dynamics of protein complexes and complex-drug interactions are difficult to characterize (and predict). In part, this contributes to the well known *in vitro*/*in vivo* discrepancy between predicted and actual drug action and efficacy. Primarily, this is due to current limitations of *in vivo* validation processes. Experimental mapping of the dynamic complexome in normal and disease states would add significantly to overcoming this obstacle. Much more rational design and screening methods are important for developing safe and effective drugs that specifically target complexes.

Emerging Opportunities

Recent progress in the development of tools and methods to map dynamic protein-protein interactions provide a mechanism through which disease pathogenesis can be better understood and new drugs can be designed. Specific challenges must be overcome for these possibilities to become reality.

Potential CF Investments

- CF investments in the following would have a transformative impact on the identification of new drugs, the functional annotation of existing drugs, and the identification and testing of candidates for polypharmacologic approaches:
- Experimental mapping of the dynamic complexome in normal and disease states
- Development of computational tools and algorithms that allow predictive models for protein-protein and protein-drug interaction to be established and tested
- Development of drugs that target dynamic protein complexes through rational drug design and through screening approaches
- Development of novel methods to structurally characterize spatially and temporally dynamic complexes

Potential Impact

Investment in this area would benefit basic and applied studies. Identification of new functionally distinct complexes that define cellular pathways will increase knowledge about pathways and signaling mechanisms shared among diseases and conditions. Mapping the complexome may have as much potential for distinguishing disease states as does mapping the genome. Progress in this area will also likely yield small molecules as probes for clinical samples and tissue engineering models. Clinically relevant impact also includes the identification of new drugs that are specific, effective, and which have better side-effect profiles than most currently used therapeutics.

Va. Bringing Difficult Structures into Reach

Challenges and Obstacles

Despite ongoing CF efforts to develop new technologies and a better understanding of the structural biology of membrane proteins, many other proteins of intense biological interest (e.g., large proteins, multi-subunit proteins, glycosylated proteins, complexes of proteins, conformationally mobile proteins and transient interactions of proteins) remain intractable to structural biological investigation.

Emerging Opportunities

Various opportunities and techniques appear ripe for investment as they are "stuck" at the inflection point of progress over time ("**B**" in Fig. 1, page 30). These include: i) small-angle X-ray scattering in solution (which requires experimental validation); ii) single-particle X-ray analysis (which needs engineering and refinement); iii) tomography (which needs improvement in resolution); and iv) powder and fiber diffraction (which need software and education).

Potential CF Investments include workshops to define and understand the current limits of emerging technologies, prioritize those for development, and improve access to these techniques as well as education on how to use them. Input from these workshops could inform the development of RFAs for methods- and software development and testing.

Potential Impact

Advances in protein structure determination will provide greater availability of biologically relevant protein structures and complexes across diseases and ICs. Expanding the protein structure universe will also yield new templates for drug design, as well as three-dimensional maps for understanding protein function and mapping genomic variation.

Note: NIGMS does invest fairly heavily in this area, through its Protein Structure Initiative;[43] however, it is likely that improved synergy and integration among investigators from various fields would be helpful.

VI. Single-Cell Analysis

Challenges and Obstacles

Population heterogeneity among cells in a given tissue is a critical issue whose importance bridges many areas of biomedicine: cancer, infectious disease, developmental processes, organs, and immune responses. However, it is well-known that current approaches are quite limited in that they can only achieve approximate ensemble analyses of cell populations. Roadblocks to progress in this area are biological and technological: Molecular and systems level description (and quantitation) of cells, organs, and disease processes requires a greater understanding of the behaviors of individual cells and the overall composition of the population.

Emerging Opportunities

Advances in engineering and nanotechnology provide the opportunity for transformative methods in single-cell and population-based analyses. The need for ultra-sensitive analytical methods and

[43] http://www.nigms.nih.gov/Research/FeaturedPrograms/PSI/

sophisticated computational tools calls for expertise from physicists, engineers, and computer scientists. It is possible that existing theorems on organizational behavior could be re-purposed for single-cell studies.

Potential CF investments in this area would go beyond most of the current emphasis on microscopic and imaging techniques (although those approaches are also useful and necessary). Potential new investments could be in mapping a single cell's epigenome, proteome, and metabolome. In addition, CF investment is needed to extend recent proof-of-principle work in single-cell genome sequencing and transcriptomics that is highly innovative, but low-throughput and far from practice. CF investments should emphasize approaches that capture living (or recently living) cells in vivo without need for overexpression or artificial constructs.

Potential Impact

The ultimate motivation for more research in single-cell analysis is the potential for *in vivo* application to disease. Developing a robust set of tools to assess (and ultimately manipulate) single cells *in situ* is a key step toward achieving that goal. This achievement would have broad applicability across biomedicine: both for basic studies and for clinical use.

VIa. Cross-Cutting Issues in Computation and Informatics

Challenges and Obstacles

One common thread of nearly all the topics discussed at the Innovation Brainstorm is data overload. In particular, there is an urgent need for integration of data sets, approaches, as well as of inquiry that addresses multiple states of health and disease. Improved data sharing, as well as access to secondary data sets, is paramount to progress.

Emerging Opportunities

More interdisciplinary opportunities are required to tackle the data challenges in biomedicine, and as such, all efforts to ease these interactions would be well-spent.

One example of an underused opportunity is cloud computing, which provides shared computational resources on demand via a computer network. This approach could be broadened within the biomedical realm, although some fields (e.g. protein folding) have already implemented it. Since data users submit tasks without possessing the software or hardware locally, the approach promotes cost and labor efficiencies.

Developing new tools and opportunities for multi-disciplinary interactions will help integrate genomic and phenotypic data sets as well as advance the study and understanding of the broadly based "environment."

Potential CF Investments

Currently, NIH-supported resources in this area are helpful but not sufficiently broad and/or powerful enough to address the growing need to integrate multiple data sets. The NIH could "democratize" this area of research by:

- Creating innovation centers
- Lowering the entry barrier for quantitative scientists
- Presenting "prediction challenge" data sets for teams to solve
- Developing and/or hosting a software commons for existing computational biologists
- Increasing the usability of web-based computational tools for biologists
- Creating an Office of Cyberinfrastructure within the CF for trans-NIH oversight

Potential Impact

Developing and sharing broad-based computational tools and making them freely available to the scientific community has the potential to vastly increase the interoperability of data sets currently being generated in 'omics studies. Doing so is necessary for full integration of knowledge that can apply across ICs and disciplines.

Session Discussion

It is important to remember the people behind these studies: Science is a human endeavor, and creativity stems from individual bright minds. However, data management and analysis problems are often too large for individuals to solve alone; thus teaming people with multiple areas of expertise is important. Another key thread of discussion about data management is open access and standardization, both of which are central to innovation and application. Finally, there is also a need for innovative thinking in which study designs have the "right" amount of data: not too little or too much to solve a well-positioned question.

A key concern at the NIH is the workforce of biomedical scientists fluent in mathematics/computer science/physics, and vice-versa. Many training programs exist in and out of the NIH funding stream, but a key bottleneck appears to be providing vehicles for self-assembly and subsequent collaboration between scientists of different ilk.

Note: the NIH recently acquired permission to grant prizes as a mechanism to attract interest to various problems of critical importance. NIH leadership is considering how and whether such approaches may advance progress on challenging problems in biomedicine.

VII. NIH Award Strategies

A common theme during the online discussion prior to the meeting and at the meeting itself was how to bring together disparate fields of science. Despite recent CF efforts and programs across the NIH, the formation of teams and integration of multiple disciplines remains a major barrier.

Potential CF Investments include:

- Student Training: Create pilot Ph.D. programs in emerging, cross-disciplinary areas (e.g., stem cell biology and bioengineering). Such a program should receive CF investment for 10 years, to develop a curriculum and engage existing faculty. Students would perform rotations in cross-disciplinary labs.
- Postdoctoral and Faculty Training Fellowships: Postdoctoral fellowship awards could be targeted to physical scientists and engineers, and trainees would be co-advised and trained in the different fields. Another idea is a faculty fellowship program that would fund salary and supplies for periods of three to nine months in which mid-career scientists and engineers would work in clinical labs or vice-versa.
- Facilitating Ph.D. and M.D. Interactions: Topical workshops could bring together clinicians and professional scientists and engineers. One idea, modeled after the National Academies Keck Futures Initiative[44], is to host a mini-course, in which attendees work in teams (8-10 people) to solve grand challenges, and then present and summarize their results. Workshop participants could also apply for seed funding afterward with awards of $100,000 and $250,000 for two years. If successful after the pilot period, awardee teams could then compete for a much larger pool of funds, around $1.5 million for five years.
- Creating Mechanisms for Small Team Proposals: Fund small, interdisciplinary team proposals (of three to four investigators partnered with companies) that are larger than multi-PI projects but smaller than Center awards. Another idea is funding a DARPA-like mechanism in which program officers facilitate team building and which allows innovative proposals that might not have fared well in traditional study sections to get done.
- Low Hanging Fruit for Immediate Impact: The NIH could fund an intermediate stage of funding that validates university or donor-sponsored initiatives. This would encourage real investments outside of NIH and provide a federal "stamp-of-approval" for innovative or cross-cutting ideas that emerge. Another idea is to provide funding to support start-up companies, enabling them to establish an infrastructure that is a pre-requisite for attracting other investors.

VIII. Molecular Classification of Disease

Challenges and Obstacles

Currently, "clinical syndromes" are often used to classify disease. The problem with this approach is that a given patient syndrome may contain significant heterogeneity with regard to molecular mechanisms of pathogenesis. As a result, the ability to identify pathogenic mechanisms in population studies is limited, as is the ability to quickly and efficiently identify who will benefit from therapeutic interventions. Thus, new approaches are needed for classifying patients and disease states that are more tied to the molecular basis of disease. Intermediate markers or "endophenotypes" may be helpful in this regard.

Another obstacle to translation is a general lack of willingness to challenge dogma, which can perpetuate stale thinking and practice.

Emerging Opportunities

[44] http://www.keckfutures.org/

Progress in this area promises to fill gaps between molecular characterization and patient disease states, as well as to identify heterogeneity in classical clinical syndrome classifications. Recent advances in technologies that allow comprehensive profiling of patients at the molecular level and association of these profiles with clinical data provide an opportunity to completely redefine the way we think about and understand disease. However, these capabilities need to be developed further and expanded for regular use in the clinic.

Potential CF Investments

Innovation is needed in the way in which we classify patients. Examples include:

- Expression analysis of patient samples
- Epigenetic analyses of distinct cell types
- Classification and measurement of behavioral symptoms
- Integration of different phenotypes
- Methods to measure the response to various perturbations

The NIH could establish a well-characterized, central sample database to encourage data sharing and integration. New approaches to finding "lenses" to view complex biomedical problems could include funding coherent, high risk programs, as well as considering the relevance and ability of existing networks to pursue this work (e.g., CTSAs).

Potential Impact

Molecular characterization of disease has obvious benefit across the board for diagnosis and treatment of all diseases. In addition, progress in this area would catalyze the transition from one-size-fits-all medicine to personalized medicine. Clinical trials could be done more quickly and efficiently, and the resources harbored by population studies may be better utilized.

Finally, encouraging a mandate to challenge dogma would likely introduce broader thinking that will undoubtedly open new avenues for exploration.

Session Discussion

Any efforts to speed the process of moving basic science to the clinic would be welcomed enthusiastically by everyone. One suggestion is for the NIH to relax the stepwise nature of grant administration. For example, NIH program staff could engage researchers with promising data before publication and perform a matchmaking role with clinical investigators who could begin designing and implementing studies (rather than waiting for publications to be written, reviewed, accepted, and published). NIH coordination, in these cases, would be central (and very different from the independent tracking of research projects by program staff that is currently accepted procedure).

Closing Remarks

Dr. James Anderson, DPCPSI Director, thanked the group for their time and effort and encouraged ongoing contact with the NIH Common Fund staff. The external input provided through the Innovation

Brainstorm meeting is an important component in the Common Fund strategic planning process. These topic areas, along with other scientific needs and opportunities proposed by NIH IC Directors, will be refined, prioritized by senior NIH leadership, and assigned to Working Groups for analysis and development into specific program initiatives that can be supported through the Common Fund beginning in FY 2013.

APPENDIX A: Selected CF Programs

Molecular Libraries Program: offers public sector biomedical researchers access to large-scale screening capacity to identify small molecules that can be optimized into chemical probes to advance biological discovery and drug development. The program operates through a three-step pipeline in which: i) assays for the research community and compounds from the Molecular Libraries collection are sent to screening centers and tested for activity in submitted assays; ii) active compounds, or "hits," are optimized by medicinal chemistry into in vitro probes for biology and disease research and drug development; and iii) information from assays is deposited into the open access database PubChem. A key goal is to facilitate connections among researchers and scientific fields. To date, the program has produced innovative chemical tools for research and drug discovery; 235 probes have been identified since 2005. Two exciting prospects include a muscarinic cholinergic receptor subtype-specific ligand that targets acetylcholine-induced vasodilation in cerebrovascular diseases or acute ischemic stroke[45]; and antagonists for the orphan receptor GPR35 that serve as novel tools to target pathways underlying pain and addiction[46].

Structural Biology of Membrane Proteins Program: aims to overcome hurdles to membrane protein isolation and three-dimensional structure determination to enhance research and drug development. It is no secret to the research community that solving the protein structures of membrane-embedded proteins has been a ferociously difficult problem. Yet, the problem is important, since 40 percent of drugs currently on the market target membrane proteins. This project, which synergizes with the Molecular Libraries program, aims to bring together experienced members of this research community to accelerate discovery of new drugs that target receptor function in disease processes, and several successes have come to pass, including solutions of CXCR4, a G-protein coupled receptor important for HIV infection, cancer growth and metastasis[47]; and the dopamine-3 receptor, one of five dopamine receptor subtypes that play a role in movement, cognition, and emotion associated with schizophrenia, Parkinson's, and drug addiction[48]. Pharmaceutical company scientists are teaming up with academic researchers who have resolved three-dimensional structures of membrane-bound proteins to accelerate drug discovery. One recent success led to a new molecule being tested in a phase-1 clinical trial.

Human Microbiome Project: aims to describe and understand the roles of microorganisms that inhabit the human body. This program is well under way, and its components include: i) a reference set of

---

[45] Bridges TM, Lewis LM, Weaver CD, Lindsley CW. Discovery of the first mAChR 5 (M5) selective ligand, an M5 Positive Allosteric Modulator (PAM). Probe Reports from the NIH Molecular Libraries Program [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2010-.

[46] Heynen-Genel S, Dahl R, Shi S, Sauer M, et al. Antagonists for the Orphan Receptor GPR35. Probe Reports from the NIH Molecular Libraries Program [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2010-.

[47] Wu B, Chien EY, Mol CD, Fenalti G, et al. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science*. 2010;330:1066-71.

[48] Chien EY, Liu W, Zhao Q, Katritch V, et al. Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science*. 2010;330:1091-5.

microbial genome sequences from five body sites; ii) demonstration projects on microbial contribution to disease; iii) new technologies for isolating and sequencing individual microbes and complex populations; iv) new computational tools; v) a data analysis and coordinating center; vi) a resource repository for cultured organisms, amplified DNA, and metagenomic DNA samples; and vii) a focus on ethical, legal and social issues. To date, the project has sequenced genomes of more than 500 microbial strains; discovered more than 29,000 novel proteins encoded by the human microbiome, and launched projects on Crohn's disease, dermatitis, obesity, abdominal inflammation, acne, and undiagnosed fever. New technologies for single-genome sequencing promise to advance this effort. It is expected that as disease-specific applications are developed, this program will migrate to relevant ICs.

Genotype-Tissue Expression Resource Pilot: aims to develop a database that correlates the genotypes of many individuals with patterns of gene expression in several tissues. This project is in the early stages, beginning with a pilot to determine feasibility through processing samples from 320 donors. If successful, the project will be scaled up to process samples from approximately 1,000 donors, generating a library of expression quantitative trait loci (eQTL).

Protein Capture Reagents Program: aims to produce a community resource of tools for studying proteins, to address the need for renewable reagents that target the entire proteome and which have broad research and clinical applications. Various approaches are being pursued, including: i) producing new antigens for the research community to make protein capture reagents; ii) scaling up existing monoclonal antibody technology using transcription factors as a test case to assess issues of throughput, reliability, cost, and scalability; and iii) developing new renewable capture technologies to address issues of cost, quality, scale, and utility. The process will take several years, and will rely on empirical results to decide whether and how to scale to the whole proteome. The NIH is working closely with other organizations, such as the Human Proteome Organisation, to avoid duplication and redundancy.

National Technology Centers for Networks and Pathways Program: aims to understand dynamic, functional relationships of proteins in health and disease. Technology developments aligned with driving biological projects produce infrastructure improvements and train new researchers, in addition to yielding advances in a range of scientific areas.[49,50,51]

Interdisciplinary Research Program: explores new approaches to foster team science to tackle complex biomedical problems. This program represents a novel NIH approach to research and its administration. Nine consortia integrate training, core services, research projects, and pilot studies focused on a common scientific topic. Program directors from different ICs manage individual awards, but work together to manage the whole. The underlying goal is to foster integration across multiple academic disciplines, investigators, and universities. One recent discovery within this program found that the

[49] Fan D, Yin Z, Cheong R, Zhu FQ, et al. Subcellular-resolution delivery of a cytokine through precisely manipulated nanowires. *Nat Nanotechnol*. 2010;5:545-51.
[50] Wang Q, Chaerkady R, Wu J, Hwang HJ, et al. Mutant proteins as cancer-specific biomarkers. *Proc Natl Acad Sci U S A.* 2011;108:2444-9.
[51] Peng T, Bonamy GM, Glory-Afshar E, Rines DR, et al. Determining the distribution of probes between different subcellular locations through automated unmixing of subcellular patterns. *Proc Natl Acad Sci U S A*. 2010;107: 2944-9.

hormone adiponectin plays a role in metabolism and obesity, interacting with ceramides to reverse cell death in pancreatic and cardiac cells[52]. The program, which began at the outset of the Roadmap, will likely move its components to ICs in the near future.

High-Risk, High-Reward Research Program: looks for new approaches to support exceptionally creative scientists who propose highly innovative approaches to major contemporary challenges in biomedical research. Through a streamlined, more individualized application process, these programs encourage early-stage and new investigators (Pioneer, New Innovator, Early Independence Awards). The Early Independence Award, in particular, addresses the rising age to research independence, which is currently 42-44 years (compared to 36-38 years in 1980). This program allows exceptional young researchers to skip the post-doctoral training period and was inspired by various programs in place at institutions across the country. The NIH sought input from these institutions and the general biomedical research community, and is currently funding a pilot of approximately 10 awards. A new doctorate locates an institution willing to host them, and institutions are also free to actively recruit eligible candidates. Pending results of the pilot, the NIH may fund more awards in future years. However, it is unlikely that this model will be broadly effective; many young scientists benefit greatly from the postdoctoral training period.

[52] Holland WL, Miller RA, Wang ZV, Sun K. Receptor-mediated activation of ceramidase activity initiates the pleiotropic actions of adiponectin. *Nat Med*. 2011;17:55-63.

APPENDIX B: INVITED PARTICIPANTS

Andrea M. Armani, PhD
Assistant Professor
Fluor Early Career Chair of Engineering
Department of Chemical Engineering and Material Science
University of Southern California
3651 Watt Way, VHE 712
Los Angeles, CA 90089
Tel: 213-740-4428
Email: armani@usc.edu

Bradley Bernstein, MD, PhD
Associate Professor
Department of Pathology
Massachusetts General Hospital
Harvard Medical School
Richard B. Simches Building, CPZN 8234
185 Cambridge Street
Boston, MA 02114
Tel: 617-726-6906
Email: bernstein.bradley@mgh.harvard.edu

Timothy Cardozo, MD, PhD
Assistant Professor
Department of Pharmacology
New York University School of Medicine
550 First Avenue, MSB 497
New York, NY 10016
Tel: 212-263-6337
Email: timc@saturn.med.nyu.edu

Erin Clark, MD
Assistant Professor
Department of Obstetrics and Gynecology
Division of Maternal-Fetal Medicine
University of Utah Health Sciences
30 N. 1900 E.
Suite 28200 SOM
Salt Lake City, UT 84132
Tel: 801-581-8425
Email: erin.clark@hsc.utah.edu

Ileana Cristea, PhD
Assistant Professor
Department of Molecular Biology
Princeton University
210 Lewis Thomas Laboratory
Princeton, NJ 08544
Tel: 609-258-9417
Email: icristea@princeton.edu

Jose C. Florez, MD, PhD
Assistant Professor
Center for Human Genetic Research and Diabetes Unit
Massachusetts General Hospital
Simches Research Building- CPZN 5.250
185 Cambridge Street
Boston, MA 02114
Tel: 617-643-3308
Email: jcflorez@partners.org

Julia Gohlke, PhD
Assistant Professor
Environmental Health Sciences
University of Alabama at Birmingham
RPHB 530
1530 3rd Avenue South
Birmingham, AL 35294-0022
Tel: 205-934-7060
Email: jgohlke@uab.edu

Tawanda Gumbo, MD
Associate Professor
Internal Medicine
University of Texas Southwestern Medical School
5323 Harry Hines Boulevard
Dallas, TX 75390
Tel: 214-648-9914
Email: tawanda.gumbo@utsouthwestern.edu

Howard Hang, PhD
Assistant Professor
Laboratory of Chemical Biology and Microbial Pathogenesis
The Rockefeller University
1230 York Avenue Box 250
New York, NY 10065
Tel: 212-327-7275
Email: Howard.Hang@rockefeller.edu

Deborah T. Hung, MD, PhD
Assistant Professor, Co-Director
Microbiology & Molecular Genetics, Infectious Disease Initiative
Harvard Medical School, Broad Institute
7 Cambridge Center, NE30-2013
Cambridge, MA 02142
Tel: 617-643-3117
Email: hung@molbio.mgh.harvard.edu

Manolis Kellis, PhD
Associate Professor
Department of Computer Science
Massachusetts Institute of Technology
32 Vassar Street, 32D-530
Cambridge, MA 02139
Tel: 617-253-2419
Email: manoli@mit.edu

Ophir Klein, MD, PhD
Associate Professor
Department of Orafacial Science and Pediatrics
University of California, San Francisco
513 Parnassus Avenue, Box 0442
San Francisco, CA 94143
Tel: 415-476-4719
Email: ophir.klein@ucsf.edu

Bradley Malin, PhD
Assistant Professor
Biomedical Informatics
Vanderbilt University Medical Center
2525 West End Avenue, Suite 600
Nashville, TN 37203
Tel: 615-343-9096
Email: b.malin@vanderbilt.edu

Teri W. Odom, PhD
Associate Professor
Department of Chemistry, Department of Material Science and Engineering
Northwestern University
2145 Sheridan Road
Evanston, IL 60208
Tel: 847-491-7674
Email: todom@northwestern.edu

Erica Ollmann Saphire, PhD
Associate Professor
Department of Immunology and Microbial Science
The Scripps Research Institute
10550 North Torrey Pines Road, IMM-21
La Jolla, CA 92037
Tel: 858-784-8602
Email: erica@scripps.edu

Christine K. Payne, PhD
Assistant Professor
Departments of Chemistry & Biochemistry
Georgia Institute of Technology
901 Atlantic Drive
Atlanta, GA 30332
Tel: 404-385-3125
Email: christine.payne@chemistry.gatech.edu

Georg Seelig, PhD
Assistant Professor
Electrical Engineering and Computer Science & Engineering
University of Washington
Campus Box 352500
Seattle, WA 98195-2500
Tel: 626-864-7097
Email: gseelig@u.washington.edu

Hui Sun, PhD
Associate Professor, Department of Physiology
Howard Hughes Medical Institute, David Geffen School of Medicine
University of California, Los Angeles
UCLA Physiology, 53-140 CHS
650 Charles E. Young Drive South
Los Angeles, CA 90095
Tel: 310-206-4017
Email: hsun@mednet.ucla.edu

Eva M. Szigethy, MD, PhD
Associate Professor
Psychiatry, Medicine, Pediatrics
University of Pittsburgh School of Medicine
4401 Penn Avenue, 5FP-5115
Pittsburgh, PA 15224
Tel: 412-692-8147
Email: szigethye@upmc.edu

Desiré Tshala-Katumbay, MD, PhD
Assistant Professor & Staff Scientist
Department of Neurology, School of Medicine
Center for Research on Occupational and Environmental Toxicology
Oregon Health and Science University
Adjunct Professor of Neurology
School of Medicine
University of Kinshasa, DR Congo
3181 SW Sam Jackson Park Road
Portland, OR 97007
Tel: 503-494-0999
Email: tshalad@ohsu.edu

Angelique Whitehurst, PhD
Assistant Professor
Department of Pharmacology, Lineberger Cancer Center
School of Medicine
University of North Carolina, Chapel Hill
4093 Genetic Medicine
Chapel Hill, NC 27599-7365
Tel: 919-699-3921
Email: awhit1@med.unc.edu

Linda Wilbrecht, PhD
Assistant Professor, Principal Investigator
Department of Neurology
Ernest Gallo Clinic and Research Center
University of California, San Francisco
5858 Horton Street, Suite 200
Emeryville, CA 94608
Tel: 510-985-3560
Email: lwilbrecht@gallo.ucsf.edu

Joseph C. Wu, MD, PhD
Associate Professor
Medicine/Cardiology and Radiology
Stanford University School of Medicine
300 Pasteur Drive, Grant S14OB
Stanford, CA 94305
Tel: 650-736-2246
Email: joewu@stanford.edu

Mehmet Fatih Yanik, PhD
Associate Professor
Department of Electrical Engineering and Computer Science
Program of Computational and Systems Biology
Massachusetts Institute of Technology
Broad Institute of MIT and Harvard University
Department of Biological Engineering
77 Massachusetts Avenue, Romo 36-834
Cambridge, MA 02139
Tel: 650-814-7067
Email: yanik@mit.edu

APPENDIX C: NIH PARTICIPANTS

James M. Anderson, MD, PhD
Director
Division of Program Coordination, Planning, and Strategic Initiatives
Office of the Director
National Institutes of Health
Email: andersonjm@mail.nih.gov

Rose Aurigemma, PhD
Chief
Office of Biodefense Research Affairs Drug Development Section
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Email: rosemarie.aurigemma@nih.gov

Ravi Basavappa, PhD
Program Director
Office of Strategic Coordination
Division of Program Coordination, Planning, and Strategic Initiatives
Office of the Director
National Institutes of Health
Email: basavapr@od.nih.gov

Josephine P. Briggs, MD
Director
National Center for Complementary and Alternative Medicine
National Institutes of Health
Email: josephine.briggs@nih.gov

Robert H. Carter, MD
Deputy Director
National Institute of Arthritis and Musculoskeletal and Skin Diseases
National Institutes of Health
Email: robert.carter@nih.gov

Francis S. Collins, MD, PhD
Director
National Institutes of Health
Email: francis.collins@nih.gov

Richard Conroy, PhD
Program Director
Division of Applied Science and Technology
National Institute of Biomedical Imaging and Bioengineering
National Institutes of Health
Email: richard.conroy@nih.gov

Milton Corn, MD
Deputy Director for Research and Education
National Library of Medicine
National Institutes of Health
Email: milton.corn@nih.gov

Stephanie L. Courchesne, PhD
AAAS Science and Technology Policy Fellow
Office of Strategic Coordination
Division of Program Coordination, Planning, and Strategic Initiatives
Office of the Director
National Institutes of Health
Email: stephanie.courchesne@nih.gov

Isabel Garcia, DDS, MPH
Acting Director
National Institute of Dental and Craniofacial Research
National Institutes of Health
Email: isabel.garcia@nih.gov

Gregory G. Germino, MD
Deputy Director
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
Email: gregory.germino@nih.gov

Franziska B. Grieder, DVM, PhD
Director
Comparative Medicine
National Center for Research Resources
National Institutes of Health
Email: franziska.grieder@nih.gov

Alan E. Guttmacher, MD
Director
Eunice Kennedy Shriver National Institute of Child Health and Human Development
National Institutes of Health
Email: alan.guttmacher@nih.gov

Richard J. Hodes, MD
Director
National Institute on Aging
National Institutes of Health
Email: richard.hodes@nih.gov

Tanya Hoodbhoy, PhD
Program Director
Office of Strategic Coordination
Division of Program Coordination, Planning, and Strategic Initiatives
Office of the Director
National Institutes of Health
Email: tanyah@mail.nih.gov

Kathy Hudson, PhD
Deputy Director for Science, Outreach, and Policy
National Institutes of Health
Email: kathy.hudson@nih.gov

Scott W. Jackson, MPA
Process Manager
Office of Strategic Coordination
Division of Program Coordination, Planning, and Strategic Initiatives
Office of the Director
National Institutes of Health
Email: jacksonsc@mail.nih.gov

Robin I. Kawazoe
Deputy Director
Division of Program Coordination, Planning, and Strategic Initiatives
Office of the Director
National Institutes of Health
Email: kawazoer@mail.nih.gov

Story C. Landis, PhD
Director
National Institute of Neurological Disorders and Stroke
National Institutes of Health
Email: story.landis@nih.gov

Teri Manolio, MD, PhD
Director
Office of Population Genomics
National Human Genome Research Institute
National Institutes of Health
Email: teri.manolio@nih.gov

Ellie Murcia, MEd
Program Specialist
Office of Strategic Coordination
Division of Program Coordination, Planning, and Strategic Initiatives
Office of the Director
National Institutes of Health
Email: murciae@mail.nih.gov

Mary E. Perry, PhD
Program Director
Office of Strategic Coordination
Division of Program Coordination, Planning, and Strategic Initiatives
Office of the Director
National Institutes of Health
Email: perryma@mail.nih.gov

Roderic Pettigrew, MD, PhD
Director
National Institute of Biomedical Imaging and Bioengineering
National Institutes of Health
Email: roderic.pettigrew@nih.gov

Patricia Powell, PhD
Chief
Science Policy Branch
National Institute on Alcohol Abuse and Alcoholism
National Institutes of Health
Email: patricia.powell@nih.gov

Lita Proctor, PhD
Program Coordinator
Human Microbiome Project
National Human Genome Research Institute
National Institutes of Health
Email: lita.proctor@nih.gov

Jacqueline Roberts
Program Specialist
Office of Intramural Research
National Institutes of Health
Email: jacqueline.roberts@nih.gov

Susan B. Shurin, MD
Acting Director
National Heart, Lung, and Blood Institute
National Institutes of Health
Email: susan.shurin@nih.gov

David Shurtleff, PhD
Director
Division of Basic Neuroscience and Behavioral Research
National Institute on Drug Abuse
National Institutes of Health
Email: david.shurtleff@nih.gov

Karen Silver, MPH
Director of Outreach and Public Liaison
Office of Strategic Coordination
Division of Program Coordination, Planning, and Strategic Initiatives
Office of the Director
National Institutes of Health
Email: silverk@mail.nih.gov

Dinah S. Singer, PhD
Director, Division of Cancer Biology
National Cancer Institute
National Institutes of Health
Email: dinah.singer@nih.gov

Jeffery A. Schloss, PhD
Program Director
Technology Development
National Human Genome Research Institute
National Institutes of Health
Email: jeffery.schloss@nih.gov

Douglas M. Sheeley, ScD
Program Manager
Division of Biomedical Technology
National Center for Research Resources
National Institutes of Health
Email: douglas.sheeley@nih.gov

Lawrence A. Tabak, DDS, PhD
Principal Deputy Director
National Institutes of Health
Email: lawrence.tabak@nih.gov

Harold E. Varmus, MD
Director
National Cancer Institute
National Institutes of Health
Email: harold.varmus@nih.gov

Nora D. Volkow, MD
Director
National Institute on Drug Abuse
National Institutes of Health
Email: nora.volkow@nih.gov

Brenda Weis, PhD
Operations Team Leader
Office of Strategic Coordination
Division of Program Coordination, Planning, and Strategic Initiatives
Office of the Director
National Institutes of Health
Email: weis@mail.nih.gov

Elizabeth Wilder, PhD
Deputy Director
Office of Strategic Coordination
Division of Program Coordination, Planning, and Strategic Initiatives
Office of the Director
National Institutes of Health
Email: betsywilder@nih.gov

Richard Woychik, PhD
Deputy Director
National Institute of Environmental Health Sciences
National Institutes of Health
Email: rick.woychik@nih.gov