



Active Centers of Excellence in Genomic Science Awards

Center for Personal Dynamic Regulomes

P50 HG007735

Howard Chang

Stanford University, Palo Alto, California

Despite the rapidly increasing capacity to sequence human genomes, our incomplete ability to read and interpret the information content in genomes and epigenomes remain a central challenge. A comprehensive set of regulatory events across a genome - the regulome - is needed to make full use of genomic information, but is currently out of reach for practically all clinical applications and many biological systems. The proposed Center will develop technologies that greatly increase the sensitivity, speed, and comprehensiveness of understanding genome regulation. We will develop new technologies to interrogate the transactions between the genome and regulatory factors, such as proteins and noncoding RNAs, and integrate variations in DNA sequences and chromatin states over time and across individuals. Novel molecular engineering and biosensor strategies are deployed to encapsulate the desired complex DNA transformations into the probe system, such that the probe system can be directly used on very small human clinical samples and capture genome-wide information in one or two steps. These technologies will be applied to clinical samples and workflows in real time to exercise their robustness and reveal for the first time epigenomic dynamics of human diseases during progression and treatment. These technologies will be broadly applicable to many biomedical investigations, and the Center will disseminate the technologies via training and diverse means.

Center Web Site: *Center for Personal Dynamic Regulomes*

Center for Genomically Engineered Organs

RM1 HG008525

George M. Church

Harvard Medical School, Boston

The Center for Genomically Engineered Organs (CGEO) will combine cutting edge genomics, genome editing technology, and tissue engineering methods to develop improved models of complex tissues. These tissues will be producible in laboratories from reprogrammed or genetically modified stem or other cells, will contain multiple cell types and vasculatures representative of natural (healthy or diseased) tissues, and will be characterized deeply at a molecular level and for overall tissue architecture. Such model tissues will greatly expedite biomedical progress by providing researchers a way to conduct preliminary tests of theories about normal and disease biology quickly and inexpensively in their laboratories before they have to move on to costly and potentially invasive experiments on animals or humans. To build the capacity to generate such models, CGEO will develop methods for comprehensively analyzing tissues in situ at a molecular level, by acquiring high-throughput RNA expression, protein expression, and epigenomic data from each of the tissue's individual cells that retains information about the locations of these molecules in the cells. CGEO will develop and use these methods to characterize model tissues important to neurobiology that will be built from neurons of different types derived from human induced pluripotent stem cells and

grown into cerebral organoids. Vascularizing these organoids, and perfusing them so as to provide them nutrients and eliminate wastes, will enable them to grow into larger and more mature forms than achieved to date, and thus improve their ability to model natural tissues. In situ molecular data obtained from these neurons and organoids will be compared with data from comparable natural tissues to assess and improve their representativeness. CGEO is a collaboration of four laboratories in the Boston area with combined expertise in advanced genomic and proteomic technology, genome engineering, stem cell technology, epigenetics, super-resolution microscopy, and tissue engineering. The CGEO team comprises Professors George Church (Principal Investigator) and Chao-Ting Wu (both from Harvard Medical School), Ed Boyden (MIT), and Jennifer Lewis (Wyss Institute at Harvard).

Center Web Site: *Center for Genomically Engineered Organs*

Center for Genome Editing and Recording

RM1 HG009490

Jennifer Doudna

University of California, Berkeley

The ability to understand normal and pathologic functions of the human genome and to translate that knowledge into effective therapies depends critically on determining how encoded genetic information confers phenotype. Recent advances in DNA sequencing and bioinformatics have provided vast quantities of genomic data that, in principle, hold the keys to advances in preventive medicine and therapeutic intervention. However, realizing the promise of personalized medicine will require accurate interrogation and manipulation of DNA sequences *in situ* at a scale and level of accuracy not currently available. The Center for Genome Editing and Recording (CGER) will address these challenges by creating technologies to detect, alter and record the sequence and output of the genome in individual cells and tissues. Building on the CRISPR-Cas9 genome engineering technology harnessed from bacteria, CGER will couple the RNA-guided DNA cleavage activity of the Cas9 enzyme to strategies for enhancing DNA sequence replacement using homology-directed double-strand break repair. In parallel, CGER will conjugate Cas9 to DNA "base editing" domains to enable accurate introduction or correction of point mutations without double-stranded DNA cleavage. Using cell-based assays, CGER researchers will interrogate specific disease-associated loci in human cells to provide new biological insights and uncover new therapeutic targets. Together, these approaches will enable the creation of any desired sequence alteration at any locus with high specificity and efficiency, with profound implications for both genome science and practical therapeutic intervention. To complement this suite of genome-manipulation technologies, CGER will also develop a high-throughput pipeline for testing the functional gene expression impacts of sequence variants responsible for human disease. This pipeline will identify and illuminate the relationships between human genome sequence variations, target gene expression and interactions with other genes. Finally, CGER will create new methods for permanently recording cell state changes in DNA so that they can be read out in a single-cell RNA-seq format. This development of molecular cell recorders will focus primarily on an evolving lineage tracer that, by enabling the generation of fate maps at unprecedented resolutions, holds the promise to revolutionize studies of normal development and disease progression.

Center for Dynamic RNA Epitranscriptomes

RM1 HG008935

Chuan He

University of Chicago

RNA modifications are ubiquitous in biology and present in all classes of cellular RNAs including eukaryotic messenger and long non-coding RNA. A large fraction of mammalian mRNA/lncRNA modifications are also known to be reversible, highly dynamic, and occur in cell type and cell state dependent manner. The dynamic RNA epitranscriptomes, those involving N6-methyladenosine (m6A) in particular, are known to regulate many cellular activities including mRNA splicing, export, cytoplasmic localization, stability, translation activity, microRNA processing, immune tolerance, and to impact cellular processes including proliferation, development, circadian rhythm, and embryonic stem cell differentiation. Consider m6A in mRNA/lncRNA as an example, dedicated writers, erasers, and readers exist in human cells to orchestrate an additional layer of complex post-transcriptional gene expression regulation. Emerging new functions of RNA modifications are expected to follow, with significant implications on many aspects of human health and disease. Despite high potentials and promises, current epitranscriptome studies are significantly hampered by the lack of technologies that enable quantitative mapping of any type of mRNA/lncRNA modifications at high resolution and high sensitivity. This proposal will develop new methods that target mRNA/lncRNA modifications, such as m6A, N1-methyladenosine (m1A), 5-methylcytosine (m5C), and 2'O-methyls (Nm) for high throughput sequencing at single-base resolution and suitable for low input RNA isolated from just hundreds to thousands of cells. New bioinformatics tools will be developed in order to facilitate data analysis. The general approaches proposed can be broadly applied to sequence RNA modifications in other RNA species including more abundant ribosomal RNA, transfer RNA, snRNA, and snoRNA as well as miRNA and piRNA. We will apply the newly developed methods to obtain base-resolution maps of RNA modifications in order to associate with human diseases, and to proof-of-principle studies in neurobiology. Our proposed research will establish high-throughput, high-resolution, and high-sensitivity methods for epitranscriptome research in all biological areas.

Center Website: *Center for Dynamic RNA Epitranscriptomes*

Neuropsychiatric Genome-Scale and RDOC Individualized Domains (N-GRID)

P50 MH106933

Isaac S. Kohane

Harvard Medical School, Boston

(awarded by National Institute of Mental Health, co-funded by National Human Genome Research Institute)

As a result of the accelerated pace of development of technologies for characterizing the human genome, the rate-limiting step for large scale genomic investigation in clinical populations is now phenotyping. This is particularly the case for neuropsychiatric (NP) illness, where phenotypes are complex, biomarkers are lacking, and the primary cell types of interest are difficult to access directly. It has become apparent that both rare and common genetic variation contributes to disease risk and that this risk crosses traditional diagnostic boundaries in psychiatry. Taking advantage of a large, already-established NP biobank could dramatically accelerate progress toward understanding the cross-disorder mechanism of action of disease liability genes. This study proposes novel applications of emerging technologies in informatics and cellular neurobiology to eliminate this phenotyping bottleneck. In doing so, it will accelerate investigation of clinical and cellular phenotypes for understanding single and multilocus/polygenic associations. Aim 1: Adapt and expand one of the largest NP cellular biobanks by parsing electronic health records with gold-standard assessment of cognition and other RDoC phenotypes. Aim 2: Define the genome-wide multidimensional functional genomics (MFG) landscape in NP disease into which the transcriptomic signature (RNA-seq) of each induced neuron (IN) representing a clinically characterized individual is projected. The projection provides the mapping from molecular to phenotypic characterization and a directionality towards healthful/neurotypical states used in Aim 3. Aim 3: Develop a probabilistic model of gene expression dependencies that will predict which small molecular perturbations are likely to shift the IN transcriptomic signature in a healthful

direction in the MFG and to then update the model based on measured perturbations in the MFG. Aim 4: Select patient samples to study in greater detail for epigenetic (DNA methylation, histone marks and RNA editing) and transcriptional control particularly with regard to activity dependent changes that have been implicated in many NP diseases. Aim 5: Here we assess just how well the clinical phenotypes are informed by the genome-wide characterizations and assess which is more robust.

Center Web Site: *Neuropsychiatric Genome-Scale and RDOC Individualized Domains (N-GRID)*

Center for Cell Circuits

P50 HG006193

Aviv Regev

The Broad Institute, Cambridge, Massachusetts

Systematic reconstruction of genetic and molecular circuits in mammalian cells remains a significant, largescale and unsolved challenge in genomics. The urgency to address it is underscored by the sizeable number of GWAS-derived disease genes whose functions remain largely obscure, limiting our progress towards biological understanding and therapeutic intervention. Recent advances in probing and manipulating cellular circuits on a genomic scale open the way for the development of a systematic method for circuit reconstruction. Here, we propose a Center for Cell Circuits to develop the reagents, technologies, algorithms, protocols and strategies needed to reconstruct molecular circuits. Our preliminary studies chart an initial path towards a universal strategy, which we will fully implement by developing a broad and integrated experimental and computational toolkit. We will develop methods for comprehensive profiling, genetic perturbations and mesoscale monitoring of diverse circuit layers (Aim 1). In parallel, we will develop a computational framework to analyze profiles, derive provisional models, use them to determine targets for perturbation and monitoring, and evaluate, refine and validate circuits based on those experiments (Aim 2). We will develop, test and refine this strategy in the context of two distinct and complementary mammalian circuits. First, we will produce an integrated, multi-layer circuit of the transcriptional response to pathogens in dendritic cells (Aim 3) as an example of an acute environmental response. Second, we will reconstruct the circuit of chromatin factors and non-coding RNAs that control chromatin organization and gene expression in mouse embryonic stem cells (Aim 4) as an example of the circuitry underlying stable cell states. These detailed datasets and models will reveal general principles of circuit organization, provide a resource for scientists in these two important fields, and allow computational biologists to test and develop algorithms. We will broadly disseminate our tools and methods to the community, enabling researchers to dissect any cell circuit of interest at unprecedented detail. Our work will open the way for reconstructing cellular circuits in human disease and individuals, to improve the accuracy of both diagnosis and treatment.

Center Web Site: *Center for Cell Circuits*

Center for Photogenomics

John A. Stamatoyannopoulos

The Altius Institute

RM1 HG007743

The Center for Photogenomics will develop revolutionary technologies that enable the direct visualization and functional profiling of human regulatory regions in intact cells, and leverage the extraordinary information content of regulatory regions to pioneer novel biological and translational applications. The Center will (1) develop technologies to simultaneously visualize and localize regulatory regions on individual chromatin templates within intact cell nuclei; (2) develop approaches to enable activity-based profiling of regulatory regions; (3) pioneer structural, functional and integrative applications of photogenomics using novel super-resolution imaging techniques; and (4) lay

the foundation for translation of photogenomic techniques to the modern clinical diagnostic laboratory, and to the analysis of cells within the context of their native tissue environments. In parallel, the Center will create a strong multi-disciplinary post-doctoral program with the aim of training, mentoring, and developing a new breed of researcher with expertise in both functional genomics and advanced imaging techniques and analysis. The Center will also employ innovative outreach programs to empower diverse researchers and pre-doctoral students with photogenomic technologies and approaches.

Genomic Analysis of Network Perturbations in Human Disease

P50 HG004233

Marc Vidal

Dana-Farber Cancer Institute, Boston

Genetic differences between individuals can greatly influence their susceptibility to disease. The information originating from the Human Genome Project (HGP), including the genome sequence and its annotation, together with projects such as the HapMap and the Human Cancer Genome Project (HCGP) have greatly accelerated our ability to find genetic variants and associate genes with a wide range of human diseases. Despite these advances, linking individual genes and their variations to disease remains a daunting challenge. Even where a causal variant has been identified, the biological insight that must precede a strategy for therapeutic intervention has generally been slow in coming. The primary reason for this is that the phenotypic effects of functional sequence variants are mediated by a dynamic network of gene products and metabolites, which exhibit emergent properties that cannot be understood one gene at a time. Our central hypothesis is that both human genetic variations and pathogens such as viruses influence local and global properties of networks to induce "disease states." Therefore, we propose a general approach to understanding cellular networks based on environmental and genetic perturbations of network structure and readout of the effects using interactome mapping, proteomic analysis, and transcriptional profiling. We have chosen a defined model system with a variety of disease outcomes: viral infection. We will explore the concept that one must understand changes in complex cellular networks to fully understand the link between genotype, environment, and phenotype. We will integrate observations from network-level perturbations caused by particular viruses together with genome-wide human variation datasets for related human diseases with the goal of developing general principles for data integration and network prediction, instantiation of these in open-source software tools, and development of testable hypotheses that can be used to assess the value of our methods. Our plans to achieve these goals are summarized in the following specific aims: 1. Profile all viral-host protein-protein interactions for a group of viruses with related biological properties. 2. Profile the perturbations that viral proteins induce on the transcriptome of their host cells. 3. Combine the resulting interaction and perturbation data to derive cellular network-based models. 4. Use the developed models to interpret genome-wide genetic variations observed in human disease, 5. Integrate the bioinformatics resources developed by the various CCSG members within a Bioinformatics Core for data management and dissemination. 6. Building on existing education and outreach programs, we plan to develop a genomic and network centered educational program, with particular emphasis on providing access for underrepresented minorities to internships, workshop and scientific meetings.

Center Web Site: *Genomic Analysis of Network Perturbations in Human Disease*

Previous Centers of Excellence in Genomic Science (CEGS) Awards

Below is a list of previous Centers of Excellence in Genomic Science (CEGS) grant awards. Grant numbers link to the NIH RePORT system, where abstracts, other information about the awards and resulting publications are included.

While active CEGS have their own institutional websites, some are not maintained once the grant ends. If a website still exists, it will be linked from the project title.

P50HG002351

Center for the Study of Natural Genetic Variation
Olson, Maynard V.
University of Washington

P50HG002357

Analysis of Human Genome Using Integrated Technologies
Snyder, Michael P.
Yale University / Stanford University

P50HG002360

CEGS: Microscale Life Sciences Center
Meldrum, Deirdre R.
University of Washington / Arizona State University-Tempe Campus

P50HG002370

Center for Genomic Experimentation and Computation
Brent, Roger
VTT/MSI Molecular Sciences Institute

P50HG002568

Genomic Basis of Vertebrate Diversity
Talbot, William S. / Kingsley, David M.
Stanford University

P50HG002790

Implications of Haplotype Structure in the Human Genome
Waterman, Michael S. / Tavaré, Simon
University of Southern California

P50HG002806

Genomic Approaches to Neuronal Diversity and Plasticity
Ju, Jingyue
Columbia University Health Sciences

P50HG003170

Molecular and Genomic Imaging Center
Church, George M.
Harvard Medical School

P50HG003233

Center for the Epigenetics of Common Human Disease
Feinberg, Andrew P.
Johns Hopkins University

P50HG004071

Center for in Toto Genomic Analysis of Vertebrate Development
Bronner-Fraser, Marianne / Fraser, Scott E.
California Institute of Technology

P50HG004952

Wisconsin Center of Excellence in Genomics Science
Olivier, Michael
Medical College of Wisconsin / Texas Biomedical Research Institute

P50HG005550

Causal Transcriptional Consequences of Human Genetic Variation
Church, George M.
Harvard Medical School

P50MH090338 / P50HG006582

An Interdisciplinary Program for Systems Genomics of Complex Behaviors
Pardo-Manuel de Villena, Fernando
University of North Carolina, Chapel Hill

Last Updated: February 10, 2018