National Human Genome Research Institute

Active Centers of Excellence in Genomic Science Awards

Center for Personal Dynamic Regulomes

RM1 HG007735
Howard Chang
Stanford University, Palo Alto, California

Tens of thousands of human genomes have been sequenced, but the central challenge is their interpretation. 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 most clinical applications and biological systems. The 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 from single cells, 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 with genomic aberrations 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 the Multiplexed Assessment Of Phenotype

RM1 HG010461
Stanley Fields
University of Washington

To date, millions of human genetic variants have been found, many in the coding or regulatory sequence of genes. However, for only a tiny fraction of these variants do we understand how the expression or function of the encoded product is affected. As a consequence, the promise of sequencing human genomes to understand human phenotypes – especially the risk for many diseases with genetic components – has gone largely unfulfilled. What is needed are facile, high-throughput methods for generating libraries of human cells bearing mutant sequence elements and for assessing these libraries to determine each variant's effect on molecular and cellular phenotypes. Thus, the Center for the Multiplexed Assessment of Phenotype, based largely in the University of Washington's Department of Genome Sciences, proposes to develop highly generalizable, reproducible and scalable technologies to generate, and assess the functional impact of, variants in human genes. In the first specific aim, the Center will establish two workhorse methods of mutagenesis to produce variants: saturation editing of genes at their endogenous loci in the human genome, and in vitro generation of variant libraries that are recombined into safe harbor sites. In the second specific aim, the Center will develop approaches to explore the impact of mutations in noncoding regions on versions of genes that have been minimized – pared down to partially remove intronic sequence but still capable of providing essential activity. Further, it will develop mass spectrometry methods to analyze variation in coding sequences for its effect on protein abundance, stability, interactions, turnover and aggregation. In the third specific aim, the Center will assess variant effects on cell morphology, behavior and internal organization by using a novel, microscopy- based phenotyping technology, and on global transcription by developing a massively parallel single-cell mRNA profiling method. Center-developed technologies will be piloted on a set of human genes with disease relevance, enabling comparisons between each variant's functional effects and the effects of known pathogenic or benign variants. This effort will inform the use in the clinic of the large-scale functional data the Center's technologies will generate. Additionally, variants will be assessed under different conditions, such as in multiple cell lines, in combination with another mutation, or in the presence of a drug. The Center will also train early career experimentalists, clinical geneticists and data scientists to obtain and use large-scale functional data. This training will include internships in Center laboratories for one to three months, and apprenticeships for one to two years. These close interactions will generate medically- and biologically-relevant results and reveal the best paths for translating the vast amounts of Center-generated functional data for clinical use. Through these new technologies and their dissemination to the broader clinical community, the Center will advance the promise of the Human Genome Project by interpreting the vast landscape of human genetic variation.

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.

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.

P50HG004233
Genomic Analysis of Network Perturbations Human Disease
Vidal, Marc
Dana-Farber Cancer Institute, Boston

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: July 16, 2019