Division of Genomic Medicine Programs

Clinical Genome Resource (ClinGen): Genetic variants are increasingly being associated with phenotypes and clinical conditions. This knowledge is already being used in clinical care, particularly to accurately diagnose inherited disorders and to target drug therapy for somatic tumors. However, the aggregation of evidence and methods for assessing genes and variants is needed to inform the clinical significance of the thousands of known genes and variants associated with disease to achieve greater integration into clinical care.

The Clinical Genome Resource (ClinGen) aims to collect phenotypic and clinical information on variants across the genome, develop consensus approaches to identify clinically relevant genetic variants, and disseminate information about genomic variation to researchers and clinicians. This resource is essential for understanding the phenotypic and functional effects of genomic variation and for implementing genomics in clinical care.

Consensus Measures for Phenotypes and eXposures (The PhenX Project): Genome-wide association (GWA) studies have identified thousands of associations between genetic variants and complex human diseases, and for some diseases, such as diabetes and Crohn's disease, pooling of multiple GWA studies by meta-analysis has led to the discovery of new gene associations. However, most GWA and sequencing studies have had relatively few phenotypic and exposure measures in common. Harmonizing data across studies that have used disparate measures to collect similar information is difficult and time consuming. Development and adoption of standard phenotypic and exposure measures could facilitate the creation of larger, more comprehensive datasets with a variety of phenotype and exposure data for cross-study analysis, thus increasing statistical power and the ability to detect associations of modest effect sizes and gene-gene and gene-environment interactions. The PhenX Toolkit provides *standard* measures related to complex diseases, phenotypic traits and environmental exposures. Use of PhenX measures facilitates combining data from a variety of studies and makes it easy for investigators to expand a study design beyond the primary research focus.

Clinical Sequencing Evidence-Generating Research Program (CSER): In 2010, the National Human Genome Research Institute (NHGRI) developed and published a vision for genomics research, including the emerging application of genomic approaches in medicine (Charting a Course for Genomic Medicine, NATURE Feb 11, 2011). This strategic plan recognized the potential benefits of comprehensive genomic data that soon would be available to clinicians with the rapid deployment of new DNA sequencing instruments and methods.

The first phase of the Clinical Sequencing Exploratory Research (CSER) initiative began when NHGRI, with co-funding from the National Cancer Institute (NCI), solicited applications to : 1) leverage the Institute's long-standing experience in genomic sequencing and analysis to ease the adoption of these methods into clinical care, 2) guide the development and dissemination of best practices for the integration of clinical sequencing into clinical care, and 3) research the ethical, legal, and psychosocial implications of bringing broad genomic data into clinical decision-making including, for example, evaluation of the risks and potential benefits associated with the return of incidental findings or information on variants of uncertain effect. The CSER consortium added three more clinical sites and one Coordinating Center, and incorporated the nine projects formerly comprising the ELSI Return of Results Consortium in mid-2013. Investigators from NHGRI's Intramural ClinSeq™ project also

participated. The first phase of CSER recruited over 6,900 participants. Over 340 papers have been published, including 21 CSER-wide Working Group papers which share best practices for genomic sequencing.

In August, 2017, NHGRI, in collaboration with NCI and the National Institute on Minority Health and Health Disparities (NIMHD), awarded the second phase of CSER, now renamed the Clinical Sequencing Evidence-generating Research (CSER) Program. The second phase of CSER now includes 6 clinical sites with a focus on clinical utility and on the recruitment of ancestrally diverse and medically underserved populations, as well as the NHGRI intramural ClinSeq™ Study Coordinating Center. External collaborators, and investigators with complementary ELSI or other expertise will be eligible to participate in CSER as ancillary members.

Electronic Medical Records and Genomics (eMERGE) Network: Since 2007, the **E**lectronic **Me**dical **R**ecords and **Ge**nomics (eMERGE) Network has been funded by the National Human Genome Research Institute (NHGRI) to link existing biorepository samples to electronic medical records (EMRs) for genomic discovery and genomic medicine research. In eMERGE Phase I (2007–2011), investigators developed and applied the necessary methods and procedures to perform genome-wide association studies (GWAS) in participants with phenotypes and environmental exposures derived from EMRs. eMERGE Phase II (2011–2015) continued to conduct GWAS studies while taking on the new challenge to begin to understand the best approach to incorporate genomic information into the EMR for clinical care. To understand the impact of incorporating genomic information, each site conducted its own genetic implementation pilot study with the first network-wide effort being pharmacogenomics (eMERGE PGx). eMERGE III (2015–2019) continues research in leveraging data from large biorepositories including sequencing approximately 25,000 participants' samples. The Network is returning the sequencing results using the EMR while assessing outcomes. The Network is helping define how to incorporate the results into EMR based clinical decision support to help clinicians manage their patients' results. In addition, eMERGE III continues to study and address the ethical, legal, and social issues related to the use of EMRs for genomic discovery and genomic medicine research, such as privacy, confidentiality, and interactions with the public, as well as the return of actionable genomic results to EMRs for use in clinical care.

Genotype-Tissue Expression (GTEx) Project: The aim of the Genotype-Tissue Expression (GTEx) Project is to increase our understanding of how changes in our genes contribute to common human diseases, in order to improve health care for future generations. Launched as a two-year pilot project supported by the National Institutes of Health (NIH) Common Fund, GTEx is a resource database and associated tissue bank in which to study the relationship between genetic variation and gene expression and other molecular phenotypes in multiple reference tissues. The resource is broadly available for furthering research. In addition, the GTEx project includes an Ethical, Legal, and Social Implications (ELSI) study to explore the effectiveness of the GTEx donor consent process. This study will help ensure that the consent process and other aspects of the project effectively address the concerns and expectations of participants in the study. GTEx is a pioneering project that uses state-of-the-art protocols for obtaining and storing a large range of organs and tissues and for testing them in the lab. Until now, no project has analyzed genetic variation and expression in as many tissues in such a large population as planned for GTEx. After a successful pilot study, GTEx reached its target enrollment of 965 at the end of 2015 and is finalizing analyses of all samples.

Implementing Genomics in Practice: Pragmatic Trials Network (IGNITE PTN): The Implementing Genomics in Practice: Pragmatic Trials Network (IGNITE PTN) builds off successful genomic medicine projects in an effort to expand the implementation of genomic medicine. The network will be conducting two clinical trials in diverse, real world settings. Genetic information and clinical decision

support (CDS) will be incorporated into the electronic medical record (EMR) to advise clinicians on the assessment of risk and appropriate interventions.

Implementing Genomics in Practice (IGNITE) I Network: The IGNITE Network was created to enhance the use of genomic medicine by supporting the development of methods for incorporating genomic information into clinical care and exploration of the methods for effective implementation, diffusion and sustainability in diverse clinical settings. The Implementing Genomics in Practice (IGNITE) Network includes new or ongoing successful genomic medicine projects that incorporate genomic information into the electronic medical record (EMR) and provide clinical decision support (CDS) for implementation of appropriate interventions or clinical advice.

Population Architecture using Genomics and Epidemiology (PAGE) Consortium: Genome-wide association studies, mostly in European ancestry (EA) populations, have identified many genetic variants related to disease, highlighting the need to further explore initial findings in non-EA populations. PAGE is a consortium of U.S. studies that focuses on analyzing the relationship between genetic variants and a range of common diseases and traits. PAGE draws from up to ~100,000 study participants, including those from non-EA groups, and will examine replication, generalization, and variant discovery in non-EA individuals. Beginning in 2011, PAGE focused on studying findings in African American, Hispanic/Latino, Asian, Native Hawaiian, and Native American participants. PAGE was renewed for a second round in 2013 to develop approaches tailored for studying non-EA populations and to conduct new analyses spanning a broad range of diseases and characteristics. This phase of PAGE designed a complete study to address the incomplete information about whether the benefits or harms of genomic research apply more broadly beyond EA populations. Diversity was incorporated into every step of the study design, including selecting the study participants to be studied, designing a new genotyping technology, and analyzing the data. The results showed that increasing diversity helped discover new genomic variants, make information about existing variants more useful, and narrow the gap between EA and non-EA populations in how useful the information is. A new grant funded in June 2019 will enable PAGE to continue studying genomic influences on complex diseases in diverse populations, adding additional -omics technologies and integrating genomic sequencing data from other studies.

Undiagnosed Diseases Network (UDN): In 2008, the NIH established an intramural Undiagnosed Diseases Program (UDP) to aid individuals plagued by longstanding medical conditions that elude medical diagnosis. From 2008 until the launch of the UDN in 2015, the UDP evaluated more than 4,000 medical records, and admitted over 1,000 patients to the NIH Clinical Center for comprehensive weeklong evaluations. Building on the early successes of the UDP, the NIH extended the program into a network of clinical sites containing the intramural UDP and extramural clinical sites. These clinical sites together with a UDN Coordinating Center and other Core Laboratories comprise the Undiagnosed Diseases Network (UDN). The goals of the Network are to diagnose both rare and new diseases, to advance laboratory and clinical research by leveraging the experience and expertise of investigators, to enhance coordination and collaboration among laboratory and clinical researchers across multiple centers, and to share resulting data and approaches widely throughout the scientific community. An important outcome of the UDN will be the improved accessibility of these important diagnostic services to patients who require them.

Analysis, Visualization, and Informatics Lab-space (AnVIL): Advances in genomic technologies, as well as decreasing costs for sequencing, has enabled genomic data to become a routine component in elucidating ways to improve health. However, the rate of genomic data generation has overtaxed existing resources available for data storage, curation, and analysis, resulting in a critical need for high-

performance computing infrastructure, and specialized bioinformatics expertise. Without the proper data management and technical support, such problems will remain a bottleneck in the use of genomic data for advancing biomedical research and its implementation into clinical care.

The NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-space (AnVIL) is a scalable and interoperable resource for basic and clinical genomic research communities, that leverages a cloud-based infrastructure to democratize genomic data access, sharing and computing across large genomic, and genomic-related, datasets. The AnVIL will facilitate integration and computing on and across large datasets generated by NHGRI programs, as well as initiatives funded by National Institutes of Health (NIH), or by other agencies that support human genomics research.

The AnVIL will provide a collaborative environment, where datasets and analysis workflows can be shared within a consortium and be prepared for public release to the broad genomics community through user interfaces. The AnVIL will be tailored for both users that have limited computational expertise as well as sophisticated data scientist users.