Review

Biological Psychiatry

Convergence of Advances in Genomics, Team Science, and Repositories as Drivers of Progress in Psychiatric Genomics

Thomas Lehner, Geetha Senthil, and Anjené M. Addington

ABSTRACT

After many years of unfilled promise, psychiatric genetics has seen an unprecedented number of successes in recent years. We hypothesize that the field has reached an inflection point through a confluence of four key developments: advances in genomics; the orientation of the scientific community around large collaborative team science projects; the development of sample and data repositories; and a policy framework for sharing and accessing these resources. We discuss these domains and their effect on scientific progress and provide a perspective on why we think this is only the beginning of a new era in scientific discovery.

Keywords: Biosamples, Data Sharing, dbGaP, NIMH Repository and Genomics Resource, Psychiatric Genomics, Team Science

http://dx.doi.org/10.1016/j.biopsych.2014.01.003

The first decade of the twenty-first century has seen an unprecedented rise in positive genetic findings for complex human diseases. After languishing for decades, the advent of the parameter-free, genome-wide association paradigm has brought about the mapping of thousands of human traits to specific loci (1), and it is reasonable to believe that we are at the precipice of a golden age of gene discovery. Although most early successes in gene mapping efforts were confined to organ systems outside the central nervous system, for which the pathogenesis had been more clearly defined, genes underlying neuropsychiatric disorders have been far more elusive. In hindsight we can observe that for neuropsychiatry the last decades of the twentieth century were spent in pursuit of ephemeral genetic signals that frequently could not be replicated. However, as the field addresses the particular challenges posed by the inherent complexity of disorders of the brain, their genetic architecture is finally beginning to unravel. With a variety of approaches from genome-wide association studies (GWAS) to next-generation sequencing to Bayesian methods of integrating data across systems, recent years have seen a dramatic upswing in positive genetic findings in psychiatry (2) that are already rewriting neuroscience textbooks.

The recent successes in psychiatric genomics are the result of a confluence of several factors, some of which were set in motion decades ago, which have now led to a considerably brighter future for our ability to delineate the genetic architecture of mental disorders. We group these drivers under four domains: 1) The Human Genome Project (HGP) and accompanying advances in technology and population genetics including a shift from targeted association studies based on candidate genes to parameter-free genome-wide analytical approaches; 2) a dramatic cultural shift from single-investigator–driven research to a team science (TS) approach that is essential to acquire the number of samples, multidisciplinary input, and expertise required to exploit rapid technological and scientific advances to solve the "hard problems" in biology; 3) prescient early and ongoing investments in bio-repositories and data-repositories such as the National Institute of Mental Health (NIMH) Repository and Genomics Resource (NRGR) with its related NIMH Phenotypic and Genetic Database (NPGD) (3) as well as the National Center for Biotechnology Information Database of Genotypes and Phenotypes (dbGaP) (4); and 4) the development and implementation of a policy framework for broad data sharing and data access (5–7). For an illustrative timeline see Figure 1.

In the following we discuss these themes in turn with a focus on the biosample and data resources available in the NRGR and dbGaP. The advances described herein represent the beginning of exploitation of available scientific resources in psychiatric genomics that portend a bright future of scientific discovery.

CONCEPTUAL ADVANCES AND TECHNOLOGY

By the early 1980s, only a few human disease genes had been identified, and it was not until the successful implementation of the positional cloning paradigm later in that decade that the true era of disease gene discovery began (8). In positional cloning, genes are identified by the proximate knowledge of their position in the genome on the basis of high-resolution genetic maps and sophisticated statistical programs to determine co-segregation and co-location of genes among affected individuals [for review see (9)]. However, it quickly became clear that, for the discovery of genes associated with complex human diseases, deeper insights into the structure and workings of the human genome and the

SEE COMMENTARY ON PAGE 3

ISSN: 0006-3223

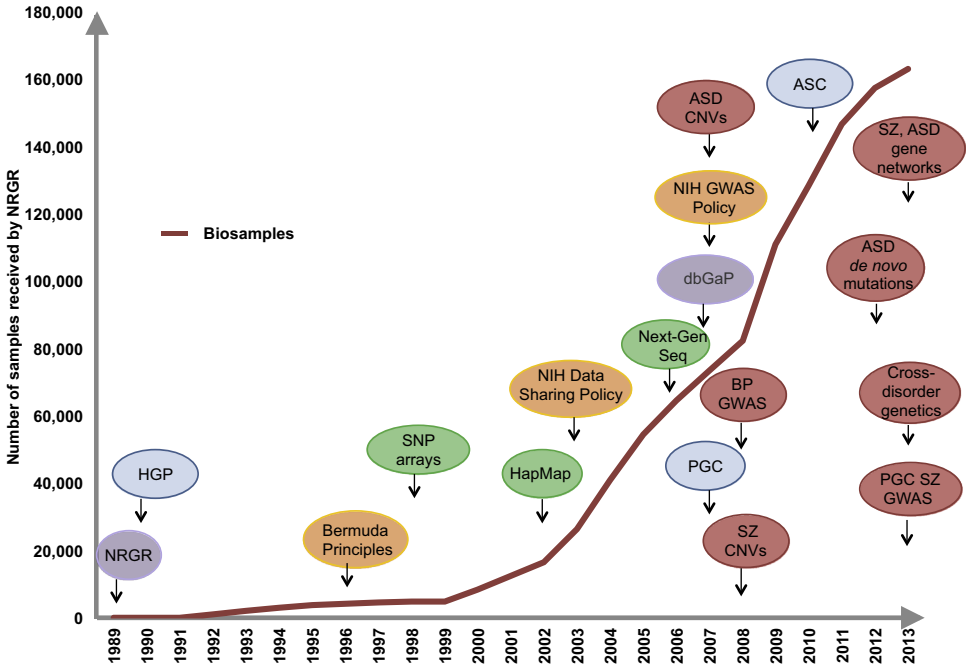

Figure 1. Illustrative timeline for progress in psychiatric genetics. Milestones that highlight the confluence of technology development and conceptual insights (green), team science (blue), data sharing policies (orange), establishment of data and sample repositories (purple), and pivotal discoveries in psychiatric genetics (red). Dark red line depicts the cumulative number of biosamples stored at the National Institute of Mental Health (NIMH) Repository and Genomics Resource (NRGR). Green: single nucleotide polymorphism (SNP) arrays, SNP arrays for genome-wide association studies (GWAS) studies (57); HapMaP, the international haplotype map, is a catalog of common genetic variants (58); Next-Gen Seq, low-cost high-throughput next-generation sequencing methods (59). Blue: Human Genome Project (HGP) (60); Psychiatric Genomics Consortium (PGC) (26); Autism Sequencing Consortium (ASC) (29). Orange: Bermuda Principles for data sharing principles and policies for public release of genome-wide data generated from HGP (5); final National Institutes of Health (NIH) statement on sharing research data (7); policy for sharing of data obtained in NIH-supported or -conducted GWAS. Purple: NRGR, established in 1989 through NIMH Human Genetics Initiative (3); the database of Genotypes and Phenotype (dbGaP) (4). Red: selected discoveries in psychiatric genetics: autism spectrum disorder (ASD) copy number variants (CNVs) (61); schizophrenia (SZ) CNVs (62); bipolar disorder (BP) GWAS (63); ASD de novo mutations (30–32); SZ, ASD gene networks (51,53); cross disorder genetics (27,64); Psychiatric Genomics Consortium (PGC) SZ GWAS (28).

fundamentals of the genetics of human populations were needed (10). The HGP (1990–2003) had the goals of providing answers to these questions and soon became the dominant force in human genetics. A decade after the completion of the draft sequence of the human genome, the HGP, together with its satellite projects (HapMap, 1000 Genomes, ENCODE [Encyclopedia of DNA Elements]) has become the main driver for advances in human genetics. Scientists have gained a plethora of insights into the complexity and diversity of the human genome, human population dynamics, and the genetic architecture of human disease (11–16). The HGP also accelerated innovation of genomic technologies and created new markets for the application of these technologies in biotechnology, bioinformatics, and healthcare. Technologies such as single nucleotide polymorphism (SNP) arrays and next-generation sequencing have allowed for increasingly cost effective large-scale characterization of genome-wide variation across genomic scales (e.g., DNA, RNA) and have become the technological drivers of the success of complex disease gene mapping in recent years. In parallel, new approaches in bioinformatics and statistics, such as high-efficiency analysis pipelines and imputation, have made possible increasingly powerful and efficient study designs and analyses of genome-wide data of neuropsychiatric disorders (17–19).

TEAM SCIENCE

The modern scientific enterprise has long acknowledged that, for many of the "hard problems" in scientific discovery, the TS approach offers many advantages, including economies of scale, resource sharing, and cross disciplinary collaboration. The HGP is generally considered to be the first successful TS project of the biological sciences and has served as an inspiration and model for other areas of research, including psychiatric genomics. A prosperous TS environment requires cultural changes that equally affect researchers, academic institutions, and funders. For example, the investigator community must learn to de-emphasize attributions by single individuals or groups and be willing to share resources across large consortia, whereas academic institutions need to develop new metrics for promotional credit that go beyond simple authorship ranking on publications. In parallel, funding agencies have to provide an environment that encourages TS. Although there are no formal policies in place, NIMH has established a number of initiatives over the last decade that encourage the TS approach and provide a supportive environment for large-scale international collaborations (20–23) with a strong emphasis on resource generation and data sharing. Furthermore, by actively engaging with the investigator community, NIMH Program staff have helped to re-orient the field toward TS, with special emphasis on "competitive collaboration," resource development, and sharing. In this context, competitive collaboration is defined as a collaborative structure that sets collaborative goals while allowing for individual groups to pursue independent research questions and outside collaborations, frequently with consortium resources. Although there is no simple metric to show the effectiveness of these strategies, a co-author network analysis of investigators