Disease Models & Mechanisms 3, 131-135 (2010)

Genomics offers new possibilities for global health through international collaboration

Tiffany Williams

Despite the biomedical advances of the last century, infectious diseases remain a leading cause of mortality and morbidity, particularly within the developing world (Fig. 1) (World Health Organization, 2008). Many of the available health technologies, including vaccines and anti-microbial drugs, fail to reach the most vulnerable, at-risk populations. Emerging and re-emerging infectious diseases exert additional pressures and strains on already struggling health systems in resource-limited settings (Seib et al., 2009). Furthermore, research funding, which is often dictated by donors from industrialized nations, does not always prioritize those non-infectious and infectious health issues of greatest importance to developing countries, the so-called '10/90' gap (Global Forum for Health Research, 2004). These and many more factors expand the existing 'North-South' health gap. The United Nation's Millennium Development Goals (MDGs) outline several target areas to redress many of the health inequities found between developed and developing nations (Fig. 2). In addition to building and strengthening health systems, science and technology innovation play a vital role to limit maternal and child mortality from infectious agents (MDG 4 and 5), and to effectively control major global pathogens (i.e. HIV/AIDS, tuberculosis and malaria, among others) (MDG 6).

As SARS (severe acute respiratory syndrome) and influenza H1N1 have shown us, infectious diseases are constrained by neither national borders nor geographic distance. Without timely and equitable distribution of information, technology and other resources, all global citizens become at-risk to these and future pathogens, independent of their nationality or location. Understanding that the diffusion of science and technology may not occur effortlessly in all directions (i.e. North to South), it is vital that policymakers, funding agencies, researchers and health care professionals promote collaborative research programs and lobby for the appropriate translation of this knowledge into practice. Experience shows that concerted efforts toward universal access to effective health technologies, such as expanded vaccination coverage in children or increased access to affordable anti-retroviral drugs, create significant gains toward global health equity. The development of simple, easy-to-use, low-cost, reliable prevention and treatment modalities that are adapted for resource-limited settings not only benefits developing countries, but may also curb the rising health care costs being experienced in developed countries.

Health technologies developed from the computational analysis of genomic and other genome-scale data possess great potential to impact our approach to control infectious diseases (World Health Organization, 2002). Genomics, the study of the structure and function of an organism's genome, and its related technologies (i.e. proteomics, transcriptomics, metabolomics) can enhance our understanding of disease pathogenesis; reveal underlying mechanisms of host susceptibility; create well-defined epidemiological disease profiles; and identify pathogen-specific

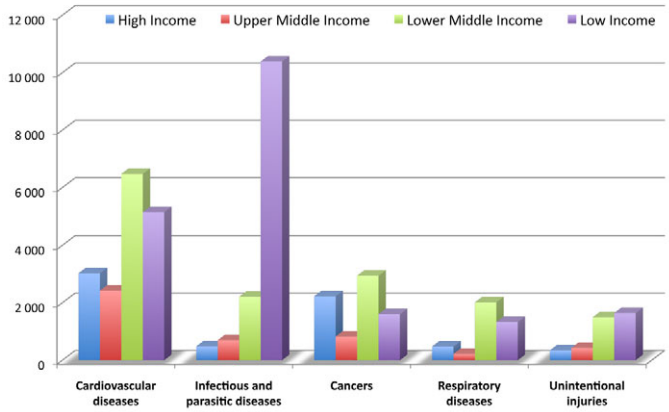

Fig. 1. The leading causes of death with respect to country income. The five leading causes of death worldwide in 2004 with respect to country income categories, as defined by the World Bank's World Development Report 2004 (World Health Organization, 2008). The y-axis shows the number of deaths ×1000. Infectious diseases ranked second, behind cardiovascular diseases, as a leading cause of global mortality in 2004. Deaths owing to infectious agents accounted for approximately 25% of all global mortality, or roughly 15 million deaths. Unlike the other top five global killers, deaths from infectious diseases were not evenly distributed across country income strata. Rather, the majority of infectious disease deaths occurred in low-income (developing) countries.

molecular targets that are suitable for further development as recombinant vaccines, therapeutics and molecular diagnostics. As described by Daar et al., three of the ten most promising biotechnologies to significantly impact global health are directly related to genomic technologies (Daar et al., 2002; Acharya et al., 2004).

In order to harness the power of genomics to control infectious diseases, it is essential to have access to genomic data, interdisciplinary collaborations and continual investment in genomic capacity, including personnel and research facilities for data analysis. As genomic knowledge is considered a 'public good', much of the data generated, as well as many of the analysis tools developed, are freely available on several public databases. Furthermore, recent advances in sequencing methodologies have resulted in the development of 'next-generation sequencing technologies' that have drastically decreased sequencing costs, while considerably increasing the rate at which data is generated and deposited in open-access databases.

Open access to genomic data and analysis tools allows researchers in developing countries to use this information to address local health issues. Using available and region-specific sequences, trained researchers in developing countries can begin to address and develop effective strategies to control infections of regional importance, with the majority of the work, apart from the actual sequencing, occurring on site. The H1N1 diagnostic test, the OptiMAL rapid malaria test, and other PCR-based techniques for the detection and classification of infectious agents are just a few examples of molecular diagnostics that have been developed through the use of genomic data with application value in developing countries. Properly adapted molecular diagnostics are often more cost-effective, sensitive and

Millennium Goals	
Goal 1	Eradicate extreme poverty and hunger
Goal 2	Achieve universal primary education
Goal 3	Promote gender equality and empower women
Goal 4	Reduce child mortality
Goal 5	Improve maternal health
Goal 6	Combat HIV/AIDS, malaria and other diseases
Goal 7	Ensure environmental sustainability
Goal 8	Global partnership for development

Fig. 2. The Millennium Development Goals (MDGs). The MDGs are eight measurable goals outlined in the Millennium Declaration, which was adopted by 189 nations in September 2000 at the UN Millennium Summit. These goals, to be achieved by 2015, seek to address the main development challenges in the world through a comprehensive and coordinated approach. The eight MDGs falls into four core thematic areas: poverty, education, health and sustainable development. Genomic applications to the control of infectious diseases can promote the achievement of MDGs 4, 5 and 6 through improved prevention, diagnosis and treatment.

specific, and less complicated to use, than non-molecular based diagnostics in resource-limited settings (Coloma and Harris, 2009). New molecular tests developed using genomic data are allowing public health officials to more effectively track disease incidence and understand the patterns of pathogen drug resistance. Recombinant vaccines developed from analyzing the genomes of the pathogens that cause meningitis, pneumonia and sepsis are in various stages of preclinical and clinical evaluation for efficacy. The use of genomic data significantly decreases the length of the discovery phase and hastens the initiation of clinical testing of new vaccines (Muzzi et al., 2007). In global health research, where resources are often limited, genomics research is a cost-effective and promising investment that will produce biotechnologies that are capable of making significant short- and long-term contributions to health.

Genomic approaches to control infectious diseases rely heavily upon the formation of interdisciplinary collaborations connecting genome-centric research centers with health professionals possessing clinical expertise in the management and treatment of these infectious agents in developing countries (Okeke and Wain, 2008). On-site researchers and clinicians are best able to articulate clinical observations, identify appropriate samples for sequencing, and translate the generated knowledge into practice. Researchers with genomic expertise are able to assess the appropriateness and applicability of various genomic tools to the problem at hand. These collaborations, when developed as equitable partnerships, increase the strength and relevance of the data to address local health concerns in developing countries. As the ultimate goal of such research is a reduction in the burden of infectious diseases, all major stakeholders, including the populations under study, must in the end be beneficiaries of any knowledge accrued. Including health care providers and researchers from developing countries in project development, implementation and analysis helps to ensure that the pathogen-specific data generated benefits the most-affected populations. Furthermore, these collaborations serve to build local genomic capacity, which contributes to the continued application of genomic technologies to health issues within developing countries.

One example of applying genomics to the problem of infectious diseases is my dissertation project at Baylor College of Medicine in the Translational Biology and Molecular Medicine graduate program. Under the direction of George Weinstock and Wendy Keitel, my work focuses on global differences in the pathogenesis of *Streptococcus pneumoniae* serotype 1 infection.

Each year, approximately one million children under the age of five, mostly living in developing countries, die from invasive *S. pneumoniae* infections. In developing countries, *S. pneumoniae* serotype 1 is commonly isolated from lethal cases of invasive pneumococcal disease (IPD) in children as well as adults. This serotype has caused meningitis outbreaks throughout Western Africa, and a significant proportion of pneumococcal-associated mortality and morbidity in developing countries can be attributed to serotype 1 infections. By contrast, within developed countries, *S. pneumoniae* serotype 1 infections rarely result in mortality and are strongly associated with pneumonia and empyema, which are milder forms of IPD, in children. The molecular epidemiology of serotype 1 shows that isolates from developing and developed countries form distinct clonal complexes. Working with colleagues at the Medical Research Council Laboratories The Gambia (MRC-TG) (Banjul, The Gambia), University of Birmingham (Birmingham, UK), Baylor College of Medicine Genome Center (Houston, TX) and the Genome Center at Washington University in Saint Louis (St Louis, MO), we analyzed how the genomic backgrounds of clinical serotype 1 isolates from developed (European and North American) and developing (African) countries influence the observed differences in virulence.

Our genomic analysis is supplemented by research with murine models of IPD. This approach allows for a direct comparison of virulence between isolates from distinct settings, with populations that have varying genetic backgrounds and biases in their basic health status. The combination of genomics and animal models, along with other molecular biology techniques, allows us to identify potential molecular targets that are likely to curb serotype 1 infections and the associated mortality through prevention, early diagnosis and effective treatment.

The molecular targets suggested by our research for therapeutic intervention may be accessed by our Gambian collaborators for their potential use as molecular diagnostics and for their further development as effective vaccines and/or therapeutics. It is common and easy to assume that many of the health problems facing developing countries are because of a lack of resources and/or access to health care. However, as our research has shown in certain cases, the pathogens themselves can contribute to the observed difference in clinical outcomes between developed and developing settings, independent of population health status. This knowledge should further help to inform policy decisions targeting serotype 1 infections within developing countries, as well as provide a data resource to begin to develop effective strategies to prevent and treat this disease.

The success of this project is testament to the synergy of our international collaboration. The field and epidemiological foundation work of our Gambian collaborators provided us with

a clear and focused scientific question, and provided access to a well-cataloged sample repository. After spending several months at the MRC-TG facilities and seeing first-hand the extensive research network that they have developed for epidemiological and clinical studies, I have a true appreciation for the pain-staking effort involved in sample and data collection by the dedicated field team at the MRC-TG. This work also highlights the absolute necessity of a 'field-to-bench' approach to translate genomic knowledge into application and evidence-based policy (i.e. 'bench to bedside'). Bolstered by our pneumococcal project and other on-going genomics-based projects, our Gambian collaborators hosted the first ever genomics-focused scientific meeting in The Gambia during December 2008. A pivotal component of this conference was a session in which local researchers and visiting scientists, including members of our pneumococcal genome project, engaged in a productive dialogue on how to build genomic capacity (including technology, personnel, computing infrastructure and the continued development of appropriate collaborations) in order to locally address the pressing health concerns of the regions.

After being introduced to this idea of using genomics to combat infectious diseases, my time as a graduate student has been a fulfilling experience demonstrating the real-world application of this concept in a feasible time scale. It is encouraging as a student to know that someone's level of training does not limit their ability to contribute to the promotion of health equity throughout the world.

Tiffany Williams (tmwillia@bcm.edu) is based at the Baylor College of Medicine, Houston, TX.

REFERENCES

Acharya, T., Kennedy, R., Daar, A. S. and Singer, P. A. (2004). Biotechnology to improve health in developing countries – a review. *Mem. Inst. Oswaldo Cruz.* **99**, 341-350.

Coloma, J. and Harris, E. (2009). Molecular genomic approaches to infectious diseases in resource-limited settings. *PLoS Med.* **6**, e1000142.

Daar, A. S., Thorsteinsdóttir, H., Martin, D. K., Smith, A. C., Nast, S. and Singer, P. A. (2002). Top 10 biotechnologies for improving health in developing countries. *Nat. Genet.* **32**, 229-232.

Global Forum for Health Research (2004). *The 10/90 report on health research 2003-2004.* Geneva: Global Forum for Health Research.

Muzzi, A., Masignani, V. and Rappuoli, R. (2007). The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterial. *Drug Discov. Today* **12**, 429-439.

Okeke, I. N. and Wain, J. (2008). Post-genomic challenges for collaborative research in infectious diseases. *Nat. Rev. Microbiol.* **6**, 858-864.

Seib, K. L., Dougan, G. and Rappuoli, R. (2009). The key role of genomics in modern vaccine and drug design for emerging infectious diseases. *PLoS Genet.* **5**, e1000612.

World Health Organization. Advisory Committee on Health Research (2002). *Genomics and world health / report of the Advisory Committee on Health Research.* Geneva: World Health Organization.

World Health Organization (2008). The global burden of disease: 2004 update. http://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/index.html

doi:10.1242/dmm.005215

SFARI Gene: an evolving database for the autism research community

Sharmila Banerjee-Basu and Alan Packer

The Simons Foundation launched its autism research initiative (SFARI; http://sfari.org) in 2003 to generate new insight into the causes of autism spectrum disorder, and to advance diagnosis and treatment. For readers of this journal, perhaps the most relevant foundation project is SFARI Gene (http://gene.sfari.org/), which will be described more fully below. This evolving database, funded by the foundation and created by Mindspec, Inc., houses comprehensive information on the human genetics of autism, as well as on relevant mouse models of the disorder.

In addition to SFARI Gene, the foundation currently funds approximately 90 principal investigators in three principal areas: gene discovery, molecular mechanisms, and cognition and behavior. SFARI's flagship project, the Simons Simplex Collection (https://sfari.org/simons-simplex-collection), engages 12 academic medical centers across North America to recruit 3000 'simplex' families with autism. Each such family has one child affected with autism, and both parents and at least one sibling are unaffected. Biospecimens are currently being banked and, upon completion of the recruitment effort in 2011, a complete database (https://sfari.org/sfari-base) will house exhaustive phenotypic information on each of the probands, as well as information generated by genome-wide surveys of copy number variation. Additional genetic data, and other data, will be imported as new research projects on the simplex collection are completed. This repository of human data is complemented by SFARI Gene, which collects valuable information from the published literature about other human genetic studies linked to corresponding animal models of autism.

Overview of SFARI Gene

It is becoming increasingly clear that autism is linked to many more genes than previously anticipated. The complex genetic architecture of neuropsychiatric disorders makes developing a genetic database for them both complicated and necessary in order to keep track of the numerous susceptibility genes uncovered by recent high-throughput methods. Several types of genetic variation, such as common variants of small effect (single nucleotide polymorphisms, SNPs), as well as rare single-gene mutations of large effect, can contribute to autism. Additionally, structural variations in the genome, such as microdeletions or duplications, are also associated with the disorder. In SFARI Gene, we have developed an integrative model and built a publicly available web portal for the ongoing collection, curation and visualization of genes linked to the disorder (Basu, 2009). The content of this resource originates entirely from the published scientific literature.

One important and unique feature of SFARI Gene is that it provides detailed annotation of the candidate genes to show their relevance to autism. All support studies are also included in the gene display page, with links to the abstract of the original articles in PubMed. Additionally, in order to provide a panoramic view of the molecular role of the autism-related genes, the reference section also includes citations that are (1) highly referred by the scientific community, and (2) recent recommendations of studies on the function of genes. A dedicated team of scientists at Mindspec continuously updates the annotations of the SFARI Gene entries with selected citations from the primary scientific literature. A panel of external advisors will soon provide additional annotations on the strength of the evidence implicating each gene in autism.

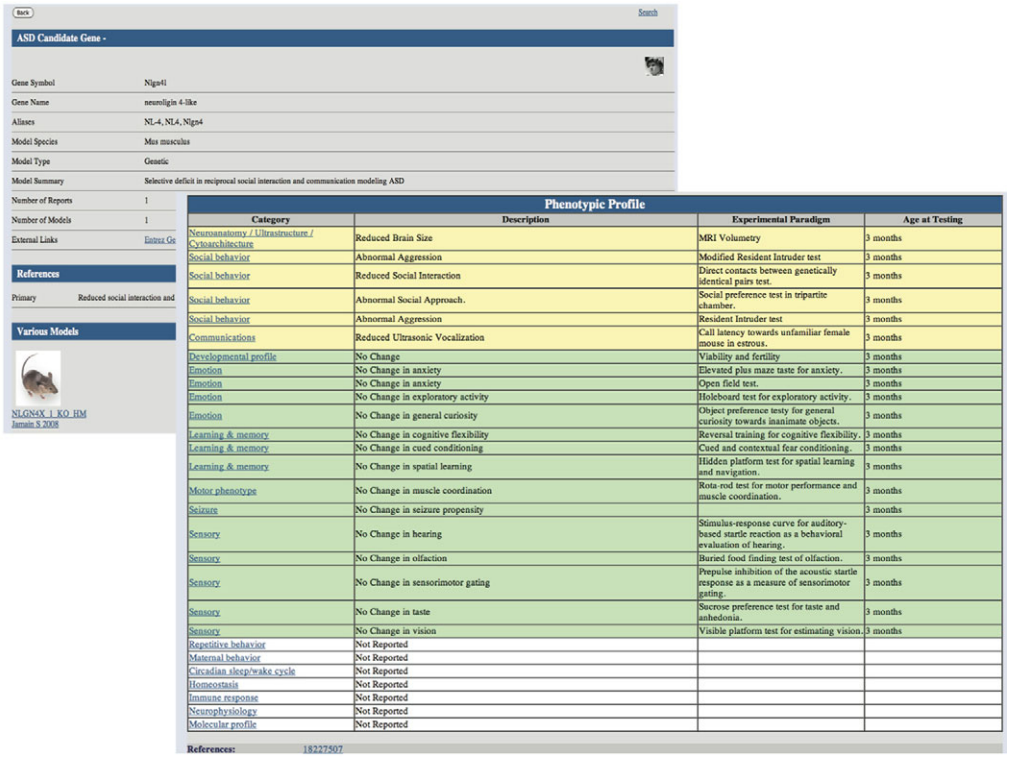

Fig. 1. Display of animal models at multiple levels. Mouse model reports pertaining to autism-candidate genes are collected by searching PubMed. Relevant information from these studies is extracted and counted for the number of reports and models. Next, the reports are collapsed under a single header representing the model gene entry. The entries are displayed at three levels: summary, model construction and phenotypic profile. The summary level includes a general description of the gene models, including the first report of the animal model(s) designated as primary reference with a link to the source abstract of the article in PubMed. Within the phenotypic profile page, an in-depth characterization of the model is provided using autism-specific annotation that has been specifically developed for this module.

A new module: animal models of autism

SFARI Gene provides a comprehensive collection of animal models linked to autism. As in the gene module of the resource, the content of the animal model module is extracted from the published scientific literature and is manually annotated by expert biologists (notably including models that were generated even before the gene in question was implicated in autism). The attributes of the animal models in SFARI Gene include the detailed description of the type of genetic construct (knockout, knock-in, knockdown, overexpression, conditional, etc.), together with the wide spectrum of phenotypic features reported in the scientific literature.

To describe the various animal models of autism in a common annotation platform, we built an additional repository of standardized terms/controlled vocabulary of mouse behavior, and other molecular features that are relevant for the biology of autism. The core behavioral features of autism involving higher order human brain functions, such as social interactions and communications, can only be approximated in animal models; therefore, our annotation strategy includes other autism-associated traits, such as seizures and circadian rhythms, which are heritable and more easily quantified in animal models. This data model attempts to capture and organize mouse phenotypic data in clinically relevant domains that are used to define autism. To this end, we developed PhenoBase, a look-up table for annotating models with controlled vocabulary, for describing animal models in a systematic fashion. One important feature of our annotation model is the inclusion of experimental paradigm in the phenotypic profile page. This feature of SFARI Gene provides a crucial assessment of the strength and specificity of the animal models. An example of an annotated animal model entry in SFARI Gene is shown in Fig. 1.

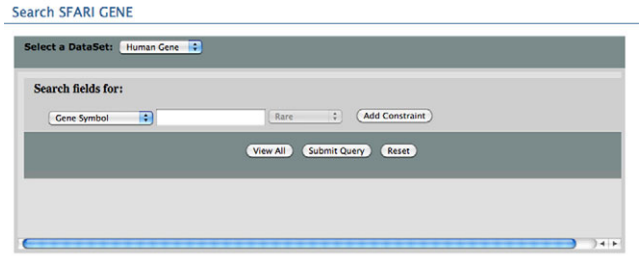

Fig. 2. The search engine in SFARI Gene. Selecting the 'animal model' dataset from the selection menu will invoke data fields that are specific for the animal model module. The animal model dataset can be searched by gene symbol, gene name, model types or phenotype categories as described in PhenoBase.

Searching SFARI Gene for animal models

The animal model module is seamlessly integrated within the gene portal so that the data can be searched and retrieved using a single search engine. This configuration essentially links two different types of datasets: genes and animal models. From the search page, users select the dataset and navigate based on their requirements (Fig. 2). The information can be searched and displayed in several ways, including complex Boolean queries. Animal model entries are displayed at three levels. At the first level of display in the summary row format, each entry is annotated with gene symbol, gene name, model species,

synteny, total number of model reports, and total number of animal models, together with a primary PubMed reference reporting the generation of the model for the candidate gene. Additionally, within the summary line display, a link is provided to the human study for the corresponding gene. Each entry is further displayed at a detail level showing (1) the gene summary, with links to external databases such as Mouse Genome Informatics (MGI; http://www.informatics.jax.org) and Allen Brain Atlas (http://www.brain-map.org/); (2) references; and (3) a list of animal models. Finally, at the third level, for each model, the extended phenotypic profile is organized under 16 categories that are relevant for the biology of autism. The animal model phenotype is shown as changes observed, no changes, or not reported (Fig. 1).

For direct participation of the scientific community, SFARI Gene offers an 'Edit' function that permits researchers to add new annotations/comments to an entry. Upon approval by a moderator, the new annotation becomes part of the entry. We view the interactive tools that accompany this modular database as essential in the effort to create a site that fully addresses the complexities of autism.

Sharmila Banerjee-Basu is the Founder/President of Mindspec, Inc. and Alan Packer is an Associate Director for Research at the Simons Foundation, New York, NY.

REFERENCE

Basu, S. N., Kollu, R. and Banerjee-Basu, S. (2009). AutDB: a gene reference resource for autism research. *Nucleic Acid Res.* **37**, D832-D836.

doi:10.1242/dmm.005439