



HHS Public Access

Author manuscript

Nat Rev Drug Discov. Author manuscript; available in PMC 2018 December 28.

Published in final edited form as:

Nat Rev Drug Discov. 2018 May ; 17(5): 301–302. doi:10.1038/nrd.2017.252.

Glimmers in illuminating the druggable genome

Griffin Rodgers¹, Christopher Austin², James Anderson³, Aaron Pawlyk⁴, Christine Colvis², Ronald Margolis⁴, and Jenna Baker⁴

¹Office of the Director, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA.

²Office of the Director, National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, Maryland 20817, USA.

³Division of Program Coordination, Planning, and Strategic Initiatives, National Institutes of Health, Bethesda, Maryland 20892, USA.

⁴Division of Diabetes, Endocrinology, and Metabolic Diseases. National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA.

Abstract

Much biomedical research continues to focus on a small proportion of the human genome that has already been studied intensively. The Illuminating the Druggable Genome programme, initiated as a pilot project by the US National Institutes of Health Common Fund in 2014, is now being implemented to accelerate the investigation of subsets of understudied proteins that have potential therapeutic relevance.

Of the ~20,000 protein-coding genes in the human genome, few have been studied intensively. Furthermore, although ~3,000 of these genes have been estimated to be part of the ‘druggable genome’ — the subset of genes encoding proteins that have the ability to bind drug-like molecules — less than 700 are currently targeted by FDA-approved drugs¹. This lack of attention to most of the proteome persists in the academic research community despite evidence that understudied proteins are often both interesting and important in biological processes², and the pharmaceutical industry tends to focus on relatively well-understood proteins as targets for drug development in order to mitigate risk. Thus, there is a substantial unmet need to both expand basic knowledge and enlarge the cache of potential

Correspondence to A.P. aaron.pawlyk@nih.gov.

Competing interests statement

The authors declare no competing interests.

FURTHER INFORMATION

GEN:<http://xt.genengnews.com/druggable-genome>

Illuminating the Druggable Genome programme: <https://commonfund.nih.gov/idg>

NIH Common Fund: <https://commonfund.nih.gov/>

Pharos:<https://pharos.nih.gov/idg/index>

Unexplored opportunities in the druggable human genome: <https://www.nature.com/nrd/posters/>

Illuminating the Druggable Genome programme homepage: <http://druggablegenome.net>

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

new drug targets through the functional characterization of understudied and poorly characterized members of druggable gene families.

Community recommendations

In 2013, the US National Institutes of Health (NIH) asked a series of expert panels, drawn from both academia and the pharmaceutical industry, to identify gaps and opportunities that could inform a new approach to studying the druggable genome. The experts identified two critical gaps: first, a lack of consolidated information on the druggable genome; and second, the need for improved technologies to characterize the function of these protein family members. They felt that filling these two gaps would enable both academic and pharmaceutical researchers to prioritize proteins more accurately for investigation as potential drug targets. They identified opportunities including the discovery of new biological activities and processes, the elucidation of the function of understudied druggable protein family members and the identification as well as the disqualification of potential therapeutic targets.

The IDG programme

The NIH Common Fund (see Further information) sponsors high-impact, multidisciplinary programmes that have the potential to transform biomedical research, are relevant to multiple diseases and conditions and cannot be achieved solely by any one NIH institute or centre. In response to the research community's recommendations, the NIH Common Fund launched a 3-year pilot programme in 2014, Illuminating the Druggable Genome (IDG; see Further information). To demonstrate the feasibility of filling the two identified critical gaps, the IDG pilot had two aims: first, consolidate disparate data types from multiple sources and make them readily available to the public; and second, adapt and scale existing technologies to assess the function of multiple understudied proteins at once. The programme focused on a few protein families with a relatively high proportion of drug targets, reflecting their prominence in physiology and medicine¹.

The pilot phase, ending in 2017, has been an unambiguous success and is currently being featured by *Genetic Engineering and Biotechnology News* (see Further information), which has partnered with IDG investigators to promote the accomplishments of the programme. Information about the proteins encoded by the druggable genome has been concentrated in a central, easily accessible location, Pharos³ (see Further information). A suite of medium- to high-throughput methods are being applied to understudied proteins across experimental systems from the molecular level to living organisms and have been made available to the greater biomedical research community (for examples, see REFS 4,5).

Implementation of the programme

Following the success of the pilot, the NIH initiated funding for a planned 6-year implementation phase of the programme. This phase establishes a research consortium to facilitate the unveiling of the functions of selected understudied IDG proteins in the druggable genome using both experimental and informatic approaches. The programme

aims to discover novel biology, with a particular focus on understudied members of the non-olfactory G protein-coupled receptor (GPCR), protein kinase and ion channel families for which there was no or minimal associated NIH funding and which had a low rate of citations in publications at the time the second funding announcement was made. It will also develop and disseminate research tools to facilitate future investigator-initiated studies since research activity on an understudied protein is often stimulated by the existence of research tools that can be used to manipulate a protein's function². The primary goal of the programme is to advance research and drug discovery through the development, broad dissemination and use of new tools and knowledge to facilitate the study of understudied human proteins. A secondary goal is to demonstrate the feasibility and benefits of illuminating their functions, permitting the expansion of such approaches to a broader array of protein families beyond the three under experimental investigation in the IDG programme.

The implementation phase of the IDG consortium will be composed of three types of centre. The first, a Knowledge Management Center (KMC), has been awarded to the University of New Mexico Health Sciences Center. Led by Tudor Oprea, it will expand the pilot phase's KMC to serve as a highly translational, protein-centric knowledge base that places data into a framework of integrated ontologies. This will be crucial for organizing and housing the resources and data produced by the IDG programme and aggregating it with existing data. Avi Ma'ayan will oversee an award to the Icahn School of Medicine at Mount Sinai that will serve as a prototype of additional informatics resources that interfaces with the KMC.

The second component is three Data and Resource Generation Centers (DRGCs). The University of North Carolina at Chapel Hill has been awarded DRGCs for understudied non-olfactory GPCRs and kinases. Bryan Roth and Brian Shoichet will lead the GPCR DRGC, which will use a combination of experimental and virtual small-molecule screening and medicinal chemistry approaches as well as CRISPR-engineered mouse lines to further probe GPCR distribution and function. Gary Johnson will explore the understudied kinome using mass spectrometric techniques applied to human cells and tissues and explore their functional role through chemical biology approaches. The third DRGC, awarded to the University of California, San Francisco, and led by Lily Jan and Michael McManus, will focus on understudied ion channels by leveraging a panel of engineered mouse embryonic stem cell lines expressing ion channel subunit constructs to explore ion channel physiology and function using proteomic and biophysical analysis, as well as mouse phenotyping.

Third, a Resource Dissemination and Outreach Center (RDOC) was awarded to the University of Miami School of Medicine and will be led by Stephan Schürer, Larry Sklar and Tudor Oprea, who will develop a system to facilitate timely public accessibility of IDG reagents and data, conduct extensive outreach to educate the scientific community about IDG efforts and coordinate the overall IDG consortium.

Goals and deliverables

The overarching goal of the IDG programme is to improve human health by defining processes to systematically illuminate understudied proteins, exploiting these processes across the proteome at the informatics level and experimentally on three exemplar families,

facilitating and tracking the uptake of these data and tools by the community and identifying new drug targets. Together with the NIH, the five centres will work with the community to define and track the most useful packages of data and toolsets to begin to illuminate the IDG proteins. Due to technical and biological reasons, the degree of illumination will vary for the selected understudied ‘dark’ proteins; however, we fully expect that the data and tools generated will provide enough information to allow multiple researchers to ‘pick up the flashlight’ and continue studying many of these understudied proteins. Not only will the programme accelerate the study of a subset of the understudied proteins, it will define a process for future systematic illumination.

Summary

The IDG programme seeks to transform basic science and drug discovery by shedding light on a subset of genes and proteins for which little publicly available information or active research exists. Elucidation of function will clear the way for proof-of-concept studies to determine the relevance of a potential therapeutic target to human health and disease. An equally important outcome will be the identification of genes and proteins that are responsible for off-target effects of existing drugs and drug candidates, as well as those proteins that are likely to be unsuitable targets for therapeutics, the knowledge of which will save valuable time and money during drug development. The long-term effect of such a resource will be expansion of the potential therapeutic space, providing an impetus for more efficient, disease-specific investigations and enhancing the ability to effectively address unmet medical needs.

References

1. Santos R et al. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* 16, 19–34 (2017). [PubMed: 27910877]
2. Edwards AM et al. Too many roads not taken. *Nature* 470, 163–165 (2011). [PubMed: 21307913]
3. Nguyen DT et al. Collating protein information to shed light on the druggable genome. *Nucleic Acids Res.* 45, D995–D1002 (2017). [PubMed: 27903890]
4. Kroeze WK et al. PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. *Nat. Struct. Mol. Biol.* 22, 362–369 (2015). [PubMed: 25895059]
5. Bruni G et al. Zebrafish behavioral profiling identifies multi-target antipsychotic-like compounds. *Nat. Chem. Biol.* 12, 559–566 (2016). [PubMed: 27239787]