

Illuminating the Druggable Genome (IDG)

**NCATS Advisory Council and
CAN Review Board Meeting**
September 15, 2016

Christine Colvis, Ph.D.
Director, New Therapeutics Program, NCATS

NCATS

Illuminating the Druggable Genome

Shedding Light on the Dark
Corners of the Genome

IDG Team

Chairs

- Christopher Austin, NCATS
- Griffin Rodgers, NIDDK

Coordinators

- Aaron Pawlyk, NIDDK
- Christine Colvis, NCATS

Project Team Leaders

- Yong Yao, NIMH (Tech Dev)
- Jean C. Zenklusen, NCI (KMC)

Project Scientists

- Ajay Pillai, NHGRI
- Margaret Sutherland, NINDS

External Scientific Panel

- Rommie Amaro, UCSD
- Andrew Hopkins, Univ. Dundee
- Peter Sorger, Harvard
- Dwight Towler, UT Southwestern

OSC

- Mary Perry
- Marishka Brown
- Aron Marquitz

Working Group Members

- Mehdi Mesri, NCI
- Ravi Ravichandran, NCI
- Matt Reilly, NIAAA
- Antonio Noronha, NIAAA
- Dwayne Lunsford, NIDCR
- Kris Bough, NIDA
- Iddil Bekirov, NIDDK
- Colin Fletcher, NHGRI
- Zorina Galis, NHLBI
- Anne Zajicek, NICHD
- Miles Fabian, NIGMS
- J. Randy Knowlton, NCI
- Enrique Michelotti, NIMH
- Corinne Silva, NIDDK
- Gurusingham Sittampalam, NCATS
- Bobbi Gardner, NCATS
- Margaret Sutherland, NINDS
- Katerina Tsilou, NICHD
- Xin (Jean) Yuan, OPA

PERSPECTIVES



OPINION

The druggable genome

Andrew L. Hopkins and Colin R. Groom

An assessment of the number of molecular targets that represent an opportunity for therapeutic intervention is crucial to the development of post-genomic research strategies within the pharmaceutical industry. Now that we know the size of the human genome, it is interesting to consider just how many molecular targets this opportunity represents. We start from the position that we understand the properties that are required for a good drug, and therefore must be able to understand what makes a good drug target.

Biological systems contain only four macromolecule with which we can use small-molecule therapeutic agents: proteins, polysaccharides, lipids and nucleic acids. Toxicity, specificity and the inability of many potent compounds against these targets means that the vast majority of drugs achieve their activity by binding to and modifying the activity of proteins. This limits the molecular targets that are commercially viable compounds.

10% drugged
3,000 genes in the druggable genome – 90% not exploited!

targets on the basis of an estimate of the number of disease-related genes¹. However, this analysis did not focus on the properties of the drugs that define those targets. The idea of assessing the number of ligand-binding domains has also recently been introduced as a measure of the number of potential points at which small-molecule therapeutic agents could act — suggestions are that this figure could be even greater than 10,000 (REF.5).

Binding sites on proteins usually exist out of functional necessity; therefore, most successful drugs achieve their activity by competing for a binding site on a protein with an endogenous small molecule. For a drug to be effective, it must bind to its molecular target with a reasonable degree of affinity. Our analysis of the *Investigational Medicines Database* (produced by Current Contents) and the *Pharmaprojects Database* (produced by PJB Publications), in addition to a thorough review of the literature, identified non-redundant molecular targets that have been shown to bind rule-of-five compounds with binding affinities that are competitive with an endogenous ligand at a structurally defined binding site.

There is some degree of overlap between the two sets of targets that we have captured several targets that are not represented by experimental data. Some targets for which no data are available have been shown to be modulated by small-molecule ligands. These targets were identified in this survey as competitive with an endogenous ligand at a structurally defined binding site.

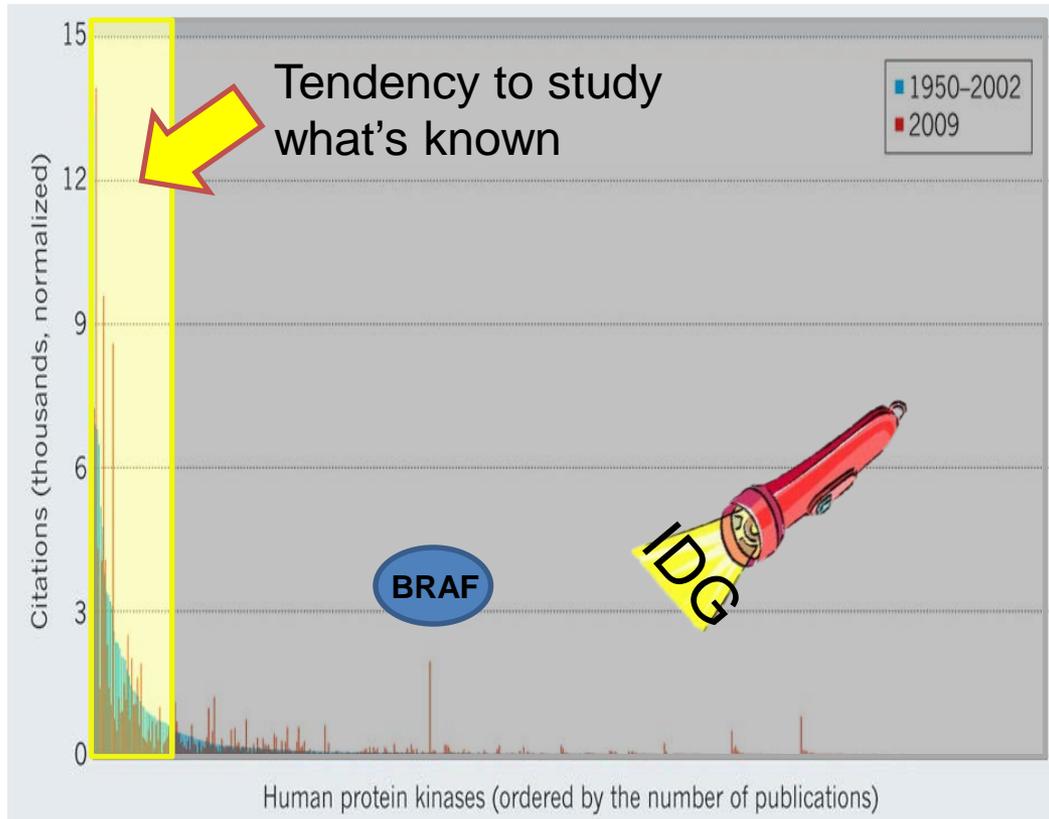
We have taken the sequences of the drug-binding domains of these proteins and determined the families that they represent, as captured by their *InterPro* domain^{6,7}. Only 130 protein families represent the known drug targets (ONLINE TABLE 1). Nearly half of the targets fall into just six gene families: G-protein-coupled receptors (GPCRs), serine/threonine and tyrosine protein kinases, zinc metallo-peptidases, serine proteases, nuclear hormone receptors and phosphodiesterases (FIG. 1a).

Box 1 | Guidelines for oral bioavailability: the 'rule-of-five'

The 'rule-of-five' analysis by Lipinski *et al.*⁸ shows that poor absorption or permeation of a compound are more likely when: there are more than five hydrogen-bond donors; the molecular mass is more than 500 Da; the lipophilicity is high (expressed as $cLogP > 5$); and the sum of nitrogen and oxygen atoms is more than 10. These rules, more appropriately described as guidelines, do not cover drugs that are derived from natural products, for which other absorption mechanisms are involved.

Clearly, published data on the oral bioavailability of existing drugs could be used as a method for defining the properties of viable drugs; however, our approach using the rule-of-five allows predictions to be made. In practice, the number of targets identified by applying the rule-of-five filters differs little from that obtained solely by literature analysis of all known drugs, whether rule-of-five compliant or not.

The IDG Challenge



Nature Reviews Drug Discovery, 2002

IDG - Pilot Phase (August 2014-August 2017)

IDG Pilot Phase Goals:

- Develop ability to identify, classify, and **prioritize understudied proteins (G-protein coupled receptors, ion channels, protein kinases, & nuclear receptors)**
- Demonstrate **scalability of technologies** needed for large-scale illumination



Knowledge Management Center (FY14-16)

- Integrate existing data and make it searchable through a single portal
- Define current state of knowledge/ignorance
- Prioritize proteins for illumination

Scalable Technology Adaptation (FY14-16)

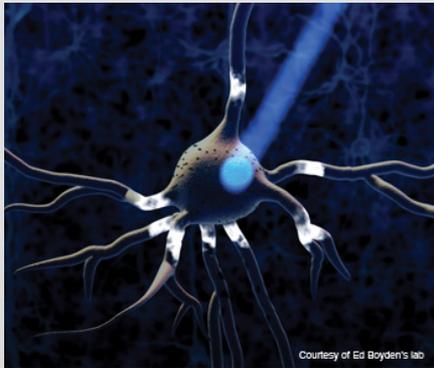
- Medium- to high-throughput, scalable assays to explore function of protein families
- New tools to facilitate experiments at needed scale

Strong interest from pharma. Seven pharmaceutical companies attended our consortium meeting in March 2016

IDG - Implementation Phase Funding Opportunity Announcements

Coming this fall

1) Expand the informatics tools developed in the pilot phase to include additional data and allow users to access a wide range of information on sets of proteins;



2) Elucidate the function of understudied proteins from key druggable protein families; and

3) Disseminate the IDG-generated resources and data to the greater scientific community.