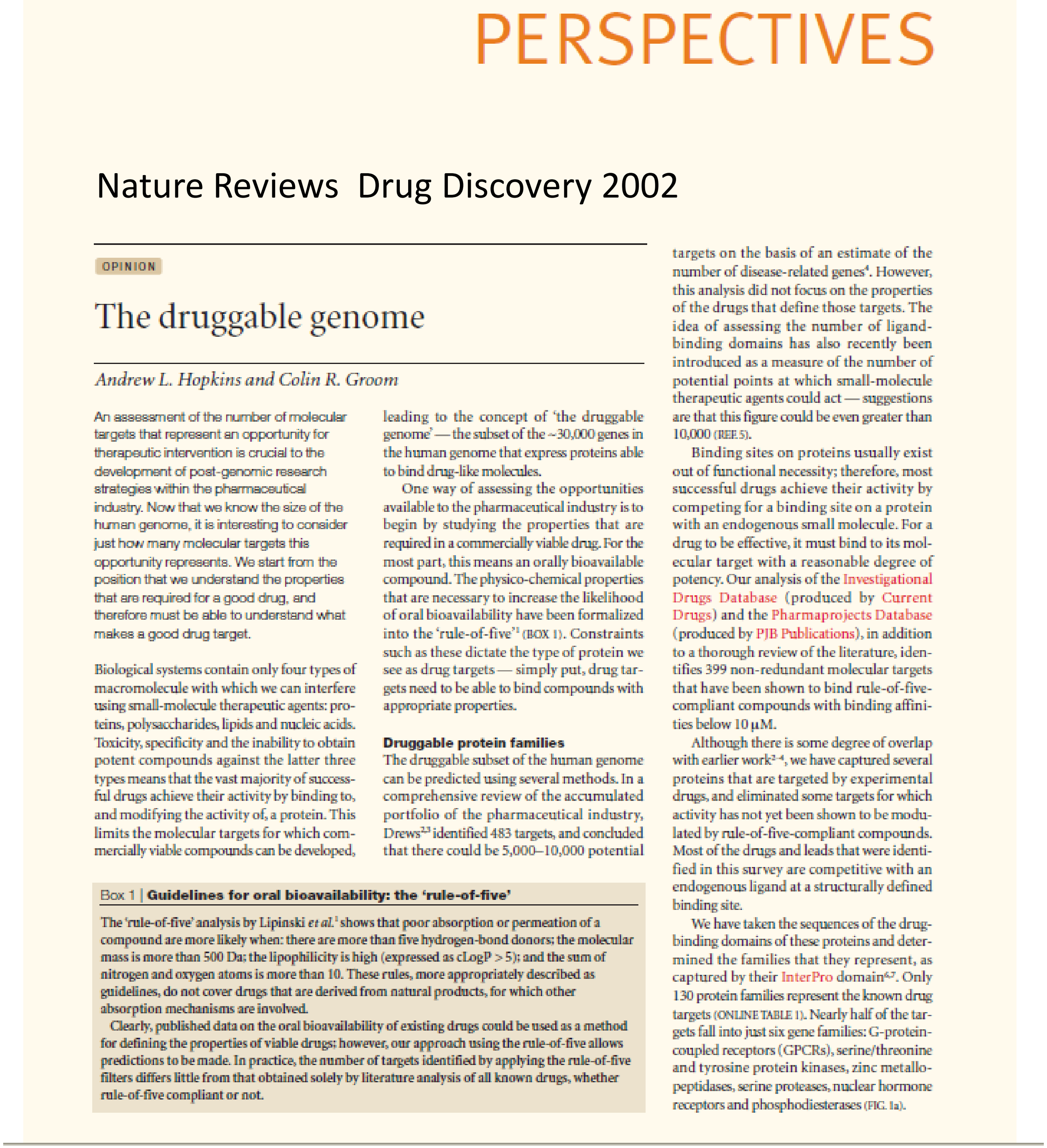
Illuminating the Druggable Genome: Common Fund proposal for FY14

Expanding the druggable genome through deep characterization of unannotated proteins



PERSPECTIVES

Nature Reviews Drug Discovery 2002

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 types of macromolecule with which we can interfere using small-molecule therapeutic agents: proteins, polysaccharides, lipids and nucleic acids. Toxicity, specificity and the inability to obtain potent compounds against the latter three types means that the vast majority of successful drugs achieve their activity by binding to, and modifying the activity of, a protein. This limits the molecular targets for which commercially viable compounds can be developed, leading to the concept of 'the druggable genome' — the subset of the ~30,000 genes in the human genome that express proteins able to bind drug-like molecules.

One way of assessing the opportunities available to the pharmaceutical industry is to begin by studying the properties that are required in a commercially viable drug. For the most part, this means an orally bioavailable compound. The physico-chemical properties that are necessary to increase the likelihood of oral bioavailability have been formalized into the 'rule-of-five'[1] (BOX 1). Constraints such as these dictate the type of protein we see as drug targets — simply put, drug targets need to be able to bind compounds with appropriate properties.

Druggable protein families

The druggable subset of the human genome can be predicted using several methods. In a comprehensive review of the accumulated portfolio of the pharmaceutical industry, Drews[2,3] identified 483 targets, and concluded that there could be 5,000–10,000 potential targets on the basis of an estimate of the number of disease-related genes[4]. 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 potency. Our analysis of the Investigational Drugs Database (produced by Current Drugs) and the Pharmaprojects Database (produced by PJB Publications), in addition to a thorough review of the literature, identifies 399 non-redundant molecular targets that have been shown to bind rule-of-five-compliant compounds with binding affinities below 10 μM.

Although there is some degree of overlap with earlier work[2–4], we have captured several proteins that are targeted by experimental drugs, and eliminated some targets for which activity has not yet been shown to be modulated by rule-of-five-compliant compounds. Most of the drugs and leads that were identified in this survey are competitive with an endogenous ligand at a structurally defined binding site.

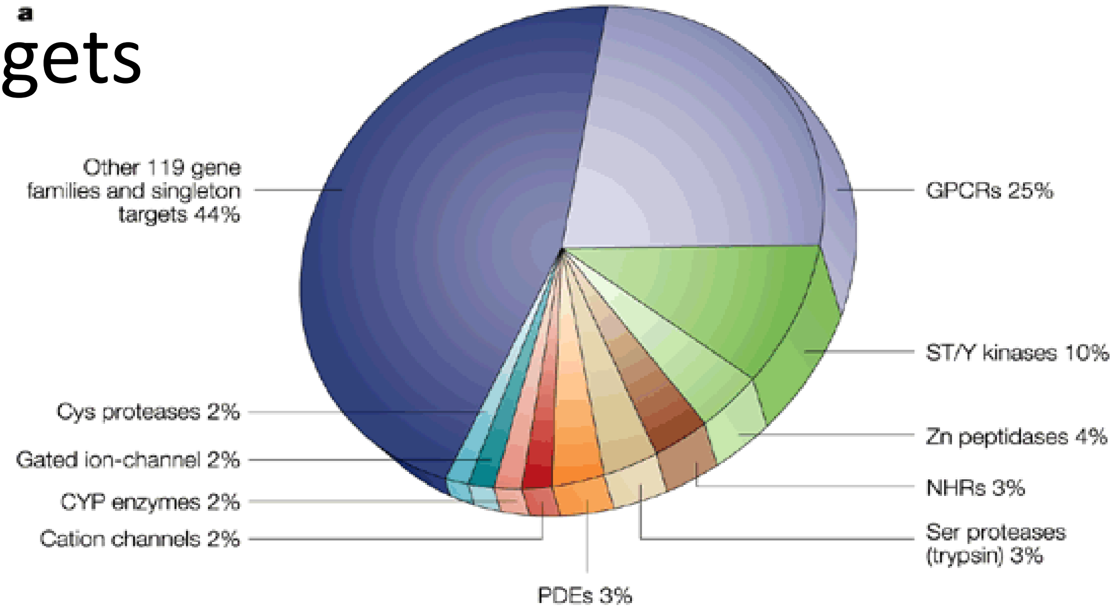
We have taken the sequences of the drug-binding domains of these proteins and determined the families that they represent, as captured by their InterPro domain[6,7]. Only 130 protein families represent the known drug targets (ONLINE TABLE 1). Nearly half of the targets fall into just six gene families: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

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

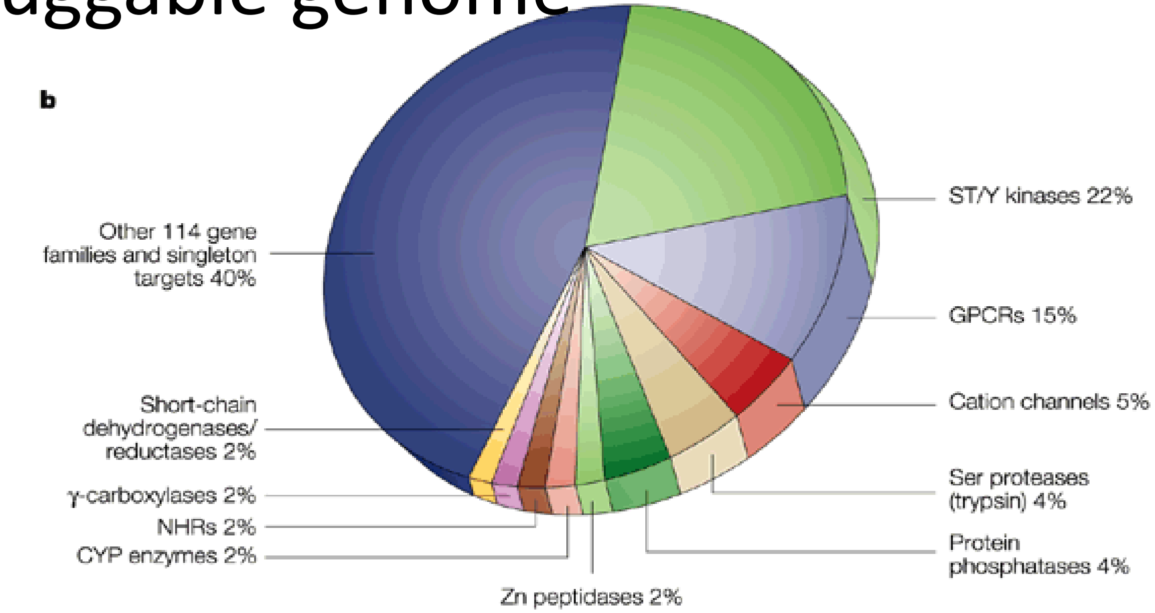
The 'rule-of-five' analysis by Lipinski *et al.*[1] shows that poor absorption or permeation of a compound are more likely when: there are more than five hydrogen-bond donors; the molecular mass is more than 500 Da; the lipophilicity is high (expressed as cLogP > 5); and the sum of nitrogen and oxygen 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.

Known drug targets



The predicted druggable genome

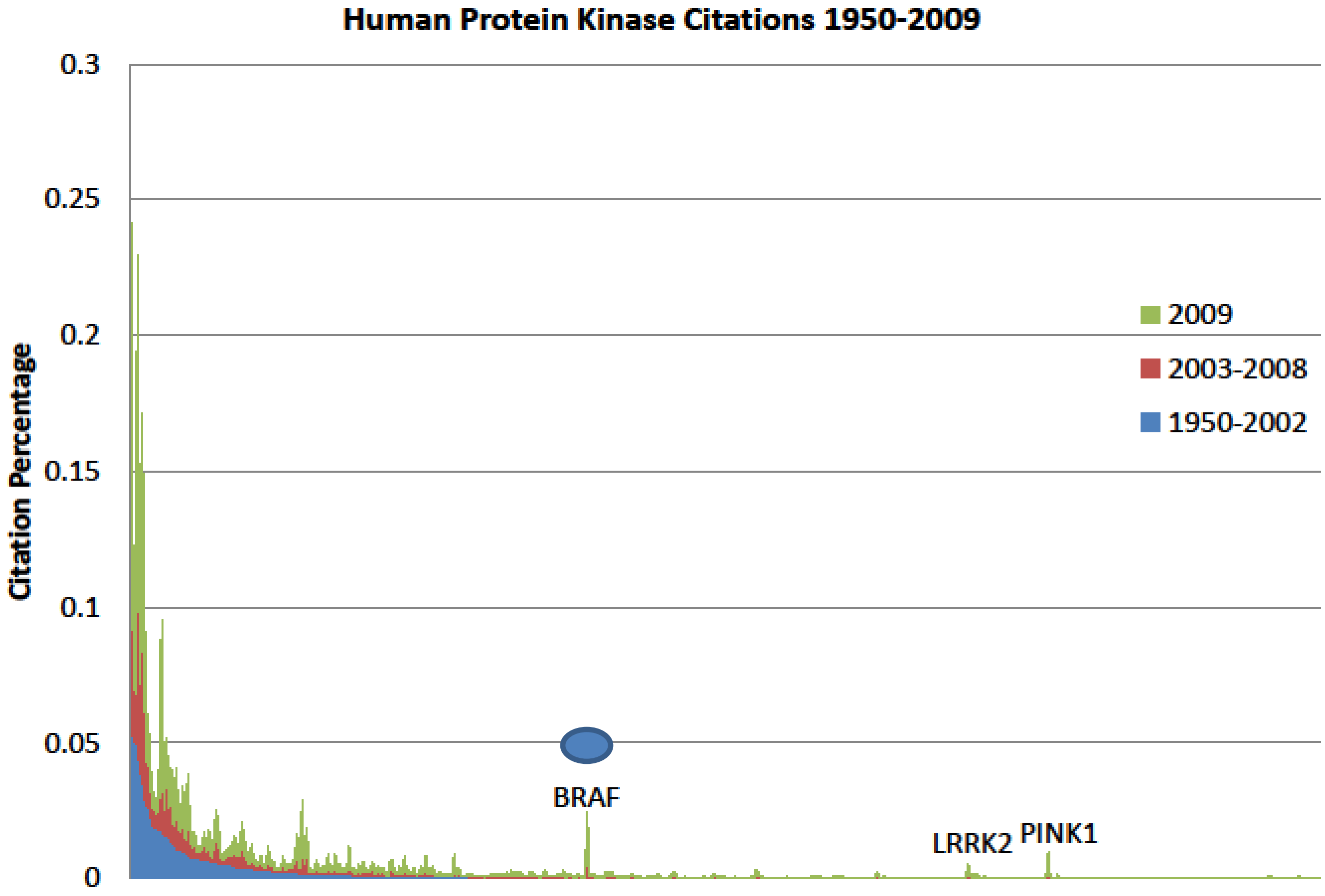


Nature Reviews | Drug Discovery

Approved CF Concept

- Deorphanize the druggable genome by:
  - Seeking understudied (unannotated) proteins within families of proteins defined as part of the druggable genome
  - Exploiting the resulting expanded druggable genome to discover new therapeutic targets

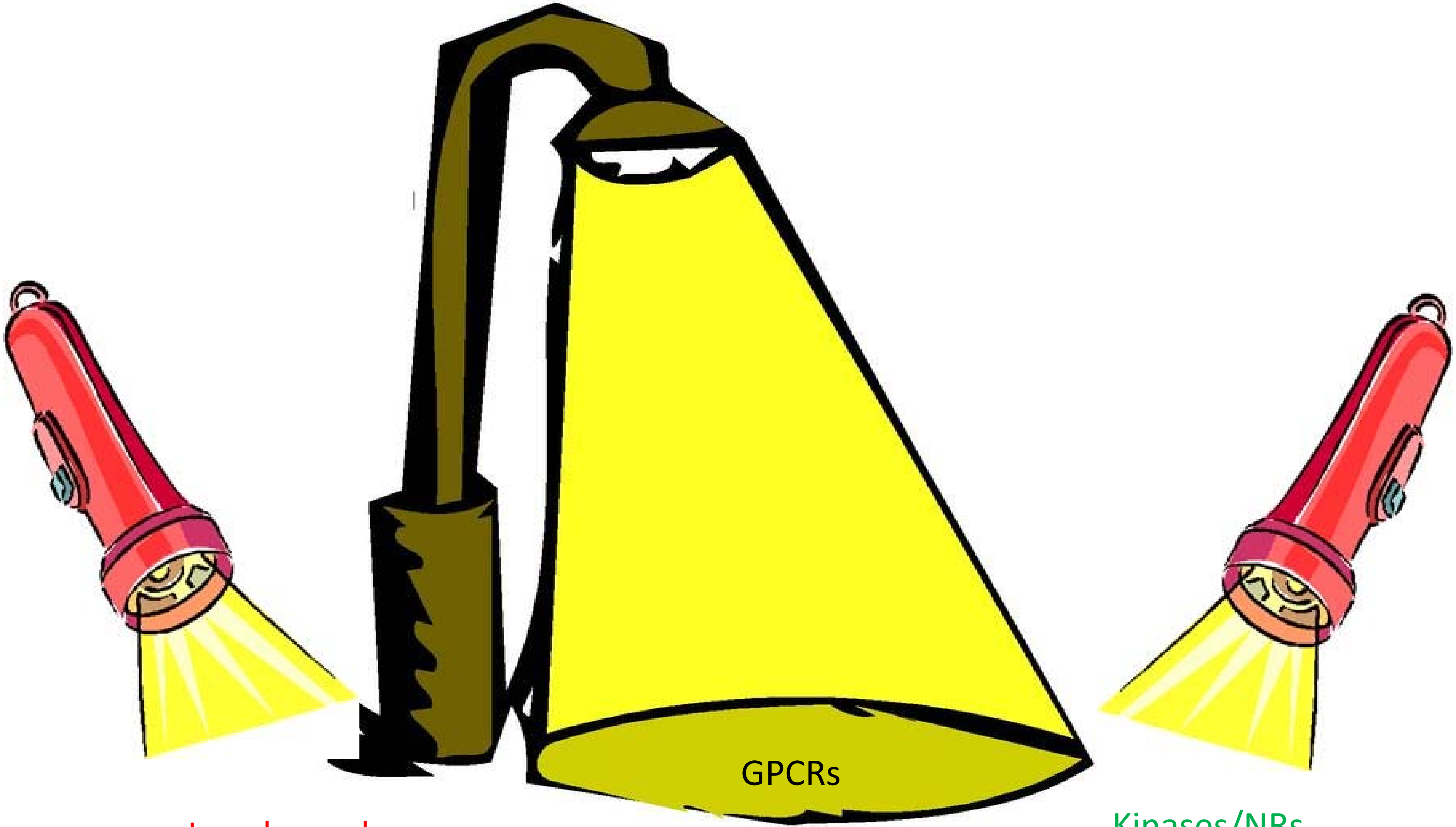
Harlow-Knapp effect: propensity to focus activities on a small fraction of the proteome



Isserlin et al. 2011 Nature

Same pattern observed for GPCRs, NRs, Ion Channels

Looking for targets under the lamppost…



GPCRs

Ion channels,
others?

Kinases/NRs

GPCRs

Gaps and Challenges

- Gaps
  - Lack of a comprehensive, curated and searchable database for the druggable genome
  - Paucity of mechanistic studies detailing how/where unannotated proteins function
  - Little understanding of the roles of unannotated proteins in disease and physiology
- Challenges
  - Pharma has abandoned many projects to expand the druggable genome
    - Mostly due to lack of knowledge of underlying biology
    - Pre-competitive space is open for exploitation by the CF
  - NIH grants system is in a Catch 22
    - Few grants awarded on unknown proteins
    - Potential of unannotated proteins is unknown

Opportunities

- Technological advances facilitate rapid and comprehensive data accrual
- Computational approaches facilitate mining large amounts of data for new information
- New, multidisciplinary fields [e.g., chemi-proteomics and chemical genomics] have created highly skilled, motivated and collaborative communities of investigators
- Synergies drawn from application of basic biology to chemistry can expand both spaces

Proposed Program for Implementation for FY14

1) Multidisciplinary research on unannotated (orphan) proteins

- Where they're expressed
- How function relates to those of other family members
- Roles in physiology and disease
- Druggability

2) Knowledge base of the druggable genome to:

- Supply consolidated, fundamental knowledge about previously unannotated proteins
- Stimulate hypothesis generation and testing, e.g., on function, polypharmacology, etc.
- Identify new potential targets relevant to unmet medical needs

3) Technology development

- New assays, targeted libraries, novel approaches, new computational tools

Outcomes

- Increased activity on the newly annotated druggable genome reflected in:
  - publications
  - new R01s
  - INDs
- Enduring publicly accessible knowledgebase
- Follow-on PPPs to take promising projects out of the CF space to exploit advances on promising targets

Deliverables

- 5 year (Phase I)
  - Expand knowledge of unannotated proteome
    - Function
    - Roles in disease/physiology
  - Develop and implement an informatics solution that creates an online, public, knowledge base of the druggable genome
    - To help identify potential drug targets relevant to unmet medical needs
    - To foster hypothesis generation and testing
- 10 year (Phase II)
  - Create pool of newly annotated potential targets
  - Determine druggability of selected proteins drawn from these pools
  - Proof-of-concept studies to determine role(s) as disease targets
  - Foster PPP to take promising projects out of the CF for development of validated therapeutic targets

Experts consulted

- Pankaj Agarwal, GSK
- Russ Altman, Stanford
- Olivier Civelli, UC Irvine
- Ron Evans, Salk Institute
- David Gerhold, NCATS IRP
- Jeff Hermes, Merck
- Andrew Hopkins, Univ. Dundee (UK)
- Bert O’Malley, Baylor COM
- John Reed, Sanford-Burnham
- Bryan Roth, Univ. North Carolina
- Peter Sorger, HMS
- Nancy Thornberry, Merck
- Dwight Towler, WUSTL

IDG NIH Working Group

- Hemin Chin, NEI
- Marciela DeGrace, NIAID
- Bonnie Dunn, NCATS
- Miles Fabian, NIGMS
- Zorina Gallis, NHLBI
- Ron Margolis, NIDDK [Chair]
- Mehdi Mesri, NCI
- Jill Morris, NINDS
- Richard Okita, NIGMS
- Aaron Pawlyk, NIDDK
- Mary Perry, OSC
- Ajay Pillai, NHGRI
- Joni Rutter, NIDA
- Belinda Seto, NIBIB
- Yong Yao, NIMH
- Xin (Jean) Yuan, OPA
- Anne Zajicek, NICHD
- Jean Claude Zenklusen, NCI

Your thoughts?