Expanding Use of Common Fund Data Sets

Ananda Roy, PhD

Program Leader, Office of Strategic Coordination

5/17/19

National Institutes of Health
Office of Strategic Coordination - The Common Fund

Enhancing the Use of Common Fund Data

Opportunity: Many CF programs generate high-value, high content, and widely available data sets.

Challenge: Many investigators are unaware of the data and/or do not have time to learn to use it. Some may want to use the data but existing computational tools are inadequate for their needs.
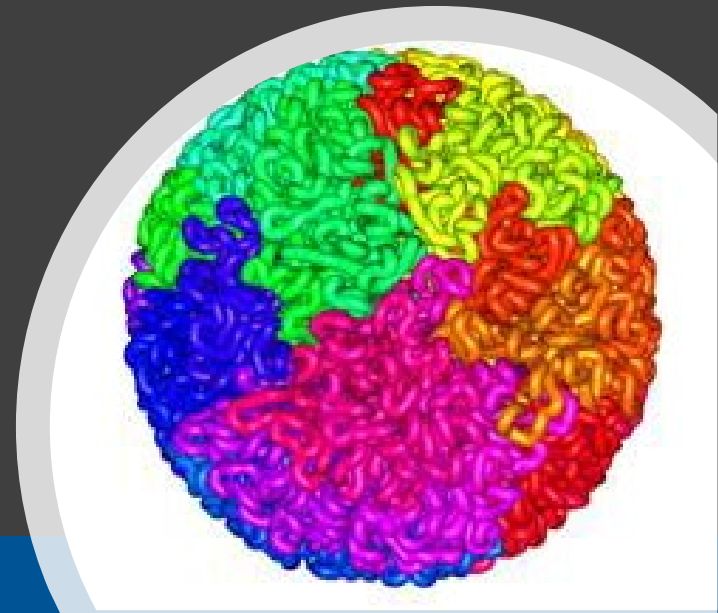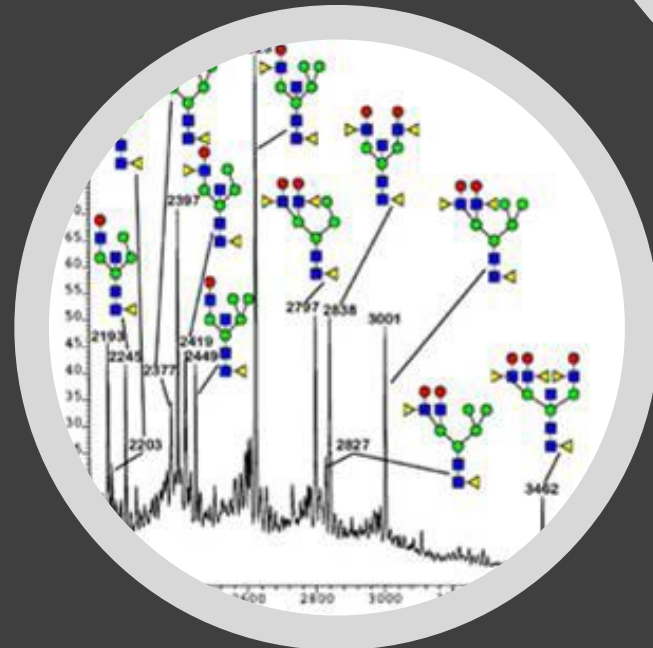
Proposal: Small projects to expand the use of Common Fund data sets

Concept:

Administrative supplements and/or other small, <u>one year projects</u> to encourage data exploration, hypothesis generation and initial testing, and/or development of novel computational tools. Projects to be solicited each year for 2-3 years as we assess impact of the initiative. Supplements open to NIH-funded PIs; small grants open to US investigators; $\leq$ $100K direct costs; expect ~10-15 awards but will depend on funds available.

Goals: To be assessed by number of users drawn to the data resources and publications that cite them

- Expand utility of Common Fund data and impact of investments

- Support investigators outside CF consortia as they learn to work with new data sets

- Support application of data and tools to a variety of research topics and diseases

Which Data Sets are Available?

- **4D Nucleome** – data on genome structure in time and space

- **Genotype-Tissue Expression (GTEx)** – genome/RNA sequence data for gene expression studies

- **Illuminating the Druggable Genome (PHAROS)** – integrated data on understudied proteins

- **Knockout Mouse Phenotyping Program (KOMP2)** – phenotype data on broad array of KO mice

- **Library of Integrated Network-based Cellular Signatures (LINCS)** – molecular signatures of cell response to disrupting agents

- **Metabolomics Workbench** – metabolomics data

- **Integrated Human Microbiome Project** – microbiome, epigenomic, metabolomic, and phenotypic data for 3 cohorts

Examples of Potential Use Cases

- Novel insights and hypothesis generation
  - ➤ *Cardiac investigators identify KOMP2 mouse mutants with heart defect; mine Kids First data to find human alleles in structural birth defect cohort; use LINCS computational tools to develop hypotheses about protein network*

- Tool development
  - ➤ *Computational scientists develop tools to enable comparison of their own data with any of the CF data sets, thereby extending the utility of the data*