Integrating functional genomics with DNA sequence variants

Mark Adams, Toby Bloom, Jim Broach, Carol Bult,
Carlos Bustamante, Deborah Colantuoni, Joe Ecker,
Elise Feingold, Kelly Frazer, Mark Gerstein,
Ross Hardison, Chanita Hughes-Halbert,
Stephen Kingsmore, Jim Lupski, Gabor Marth,
Rick Myers, Debbie Nickerson, Mike Pazin,
Len Pennacchio, Ulrike Peters, Aviv Regev,
Jay Shendure, Mike Snyder, Simona Volpi, Peter Good

Opportune time to study "function" on a large scale

- Huge number of variants available from many studies from NHGRI & beyond
  - Functional characterization = connection between genomes & biology
- Recent development of new technologies
  - CRISPR, large-scale epigenomics, single cell, etc.

We need a foundational resource to integrate functional information on many discovered variants

What should the resource be?

- Different types of function
  - At molecular/biochemical and cellular levels
    - can be studied at scale & systematized
    - Also, is closer to the variants
  - At organismal level
    - Not as easy to scale or to systematize
- NHGRI should find the “sweet spot”
  - Problems that capitalize on the new technologies
    - Lots of readout with modest investment
  - Best models – cells? mouse? model diseases?

Dichotomy of Directions

- Top-down: Develop catalogs of elements &/or all possible variants & then intersect them with variants found in disease studies
  - ex: Shendure challenge talk

- Bottom-up: Start from a list of disease variants & characterize them functionally

Both have merits

Multiple Approaches

- Approaches that look at large numbers of genes, variants, cell types, etc. in a standardized, high-throughput way

- In contrast: Deep disease/gene-specific studies
  - Require domain experts & detailed assays, many of which cannot be scaled
  - Not the province of NHGRI -- at least not on their own

- Important to have both & integrate them
  - Build special informatics infrastructure to tie them together

Other considerations

- Scaling from the genome-scale assays to population-scale
  - Success of eQTL & related projects
  - Personal functional genomics, value in longitudinal studies

- Functional genomics is valuable beyond just variant characterization
  - Use high-throughput sequencing to characterize cell types
    - e.g., to develop cellular biomarkers
    - ex: Regev challenge talk (Single-cell transcriptomics & Human Cell Atlas Project)

Integrating function & sequence variation: The opportune moment

- Large-scale resource projects as frameworks for functional characterization
  - Scaling of molecular & cellular assays v organismal phenotypes
  - "Top down" & "bottom up" both good
  - Need the interaction between standardized catalogs & domain experts

- Other aspects
  - Scaling beyond the genome to the population
  - Functional genomics regardless of variants