Morning Presentations

Summary Slides

Opportunities (1)

- Disease: focus on exemplar diseases
- Samples: encourage identification and aggregation of large, well-phenotyped, broadly consented samples
- Resource: set of recallable sequenced genomes, e.g. LoF carriers for every human gene
- Technology: continue focus on sequencing and statistical/computational methods and tools
- Whole genome sequencing: time to do more

Opportunities (2)

- Information: more active data aggregation and sharing, knowledge sharing
- Discovery and translation: a virtuous circle if we take advantage
- Functional characterization of variants: prospective, high-throughput
- Training: invest more in genome science, and statistics and computational science
- Genotyping: genotype arrays on huge samples

Summary: What has been the impact of CMG discoveries?

- CMGs have made relatively inexpensive, high-throughput gene discovery for MCs available worldwide
- Enormous amount of information about the biological function of each gene is provided by each MC “solved”
- Changing the thinking about extent of pleiotropy and genetic heterogeneity
- Enabled diagnostic and predictive testing for hundreds of MCs that that were undiagnosable
- Added 100s of starting points for the development and testing of targeted therapies
  - Key - only ~300 proteins targeted by current therapeutics

Summary: What differences have the CMGs made?

- Catalyzed the discovery of genes underlying Mendelian Conditions conditions
- 100s of new phenotypes and novel genes for Mendelian Conditions delineated
- 100s of “novel” genes for Mendelian Conditions
- Found new biological mechanisms for Mendelian Conditions
- Fostered development of statistical framework for assessing causality of variants for Mendelian Conditions
- Equipped PIs in the human genetics community with tools and skills to interpret and in many cases complete their own analysis

What are the challenges? -3

- outcomes (including cost effectiveness)
- diverse ancestries
- training
- expand scope to non-academic clinical settings
- implementation and integration in EMR environments
- mechanics of information generation and delivery; e.g. interacting with CAP
- need for large datasets linking genotypes and phenotypes
- Interfacing with regulators (e.g. FDA) and payers

If NHGRI doesn't take coordinated action, the promise of genomic medicine will be delayed

NHGRI imperatives – maximize benefit/minimize risk

- which patients, which targets.
- analysis of genomes for discovery and implementation
- accrual of Large genotype-phenotype datasets across ancestries to understand variant function
- work out the realities of implementation: consenting, EHR integration, patient and provider education, clinical decision support, follow-up…
- promoting analysis of economic and health outcomes

Breakout Session (Gerstein/Myers)
Integrating functional genomics with DNA sequence variants

- **1) What is function in genomics & how do we use it to determine the effect of variants?**
  - What are the different aspects of function and why is it hard to study? For instance, molecular (or biochemical) function vs cellular role vs organismal phenotype.
  - What are the problems in defining function? Is it meaningful to localize a function to a single place on the genome so it can be affected by a single variant? How should one think about the functional effect of large block variants?
  - Is it possible to quantitatively systematize some aspects of function so that they can be precisely related and correlated with genomic variants? In particular, what are the paradigms available to inter-relate function with variants (eg QTLs & allelic effects and phenotypes resulting from a single disruption)?

- **2) How do we inter-relate function & variants on a large scale?**
  - Is this best done by individual investigators pooling together individual results into a database or is it best done by large-scale, highly standardized experiments? What is the role of special big data database architectures for aggregating the knowledge of many functional assays?
  - Is it more effective to follow up on the many disease-associated variants uncovered by sequencing in great detail rather than doing broad genome-wide functional characterization beforehand?
  - Are there ways for new high-throughput technologies and computational approaches to significantly help with this endeavor?
  - How do we prioritize those experiments and assays that provide more functional information compared to others? Is there a particular way of assessing the information in particular experiments?

- **3) How do we validate functional effects of variants in genomics?**
  - Is it possible to validate thousands (or millions) of assertions about the genome with one or two small-scale validation experiments?
  - Is it possible to do validation at a very large scale? Is medium-scale validation possible and useful? How to think about the cost of this?
  - How do we incorporate the results of validation into quantitative error estimates for the functional assertions being made?