Known Variants Working Group

Russ Altman

Les Biesecker

David Goldstein

Heidi Rehm

Introductory comments

- Decided to focus on the questions about prior knowledge that are usually asked, and what data are available to answer them
- Some discomfort with the basic premise of known variants since in reality variants seen before are on a continuum of evidence connecting them to any given trait or constellation of traits

Documentation of the range of phenotypes that have been associated to a variant

- Important considerations for evaluating evidence for a variant

a) Evaluation be tied to a specific phenotype

b) Methodological details of phenotyping

c) Standard nosology for describing the phenotype

Resources – locus specific variant databases, OMIM, primary literature

The number and frequency of the finding of the variant in people with the phenotype (cases)

- 'Number' and 'frequency' used separately in order to capture replication of finding and its relative contribution to the total mutation load of the gene

  e.g. CFTR deltaF508 vs. a variant seen once

a) That the affected did not have another mutation in this gene
b) That the affected did not have an implicated mutation in another gene

The number and frequency of the finding of the variant in people without the phenotype (controls)

- Ethnic/genomic matching of the controls to the cases

How numbers in cases and controls compare (relative risk) and how the frequency of a mutation compares to the disease frequency

Current tools and resources to assess include:

- *1000 genomes*
- *NHLBI ESP*
- *dbSNP*
- *dbGaP*
- *Source publication for the variant descriptions*
- *Disease frequencies from GeneReviews and other publications*

Genetic data

- Segregation in multiple family members

  accompanied by proper statistical genetic analysis
- Proper allelic status

  some locus specific databases
- *De novo* germ line occurrence in sporadic disease
- Somatic evidence

  e.g. McCune-Albright syndrome and somatic GNAS mutations
- Penetrance of the variant

Mutational spectrum data

- The types of mutation in the gene
- The distribution of mutations throughout the gene (domains)

Resources: HGMD, LOVD/LDSB,dbVAR, ClinVar, ISCA, DECIPHER, OMIM, review articles including GeneReviews, and primary literature

Functional data

- Controlled experiment where mutation significantly perturbs the protein function
- Biochemical evidence of the disease in the subject is supportive though not conclusive as evidence

  e.g. α-galactosidase activity in a patient with a GLA mutation

  Contrast with an expression effect of a variant associated with depression
- Cautious of pseudodeficiencey

  e.g. Tay-Sachs

Resources: LSDB, HGMD, OMIM, local, and institutional databases

Protein prediction programs/conservation

- A tool that predicts that amino acid change is deleterious or variant has potential to change splicing
  - Such prediction judged weak guide to pathogenecity (though perhaps better guide to whether variant is deleterious, which is to be distinguished from pathogenic for a given condition)

Resources: software tools including PolyPhen, SIFT, MutationTaster, NNSplice, ESE/ESS, conservation

Pathophysiologic plausibility

- Biological evidence that perturbation of gene product is a plausible cause of the phenotype
- Includes therapeutic targets, tissue-specific expression, pathway involvement, biochemical data, etc.
- Absent when evaluating variants in uncharacterized transcripts
- Absence of biological evidence significant evidence against implication

Resources: OMIM, Unigene, GeneCards, PharmGKB, primary literature

Provenance/assessment

- Expertise, experience, reputation of the source of each type of data
- A summary or synthetic analysis of data and conclusions reached by prior evaluations of the variant
- Suboptimal practice that is currently serving as a proxy
- Long-term goal to move towards a clear and comprehensive access to data instead of using expert opinion

Key Points Slide

- Importance of integrating data – example http://matt.might.net/articles/my-sons-killer/
- What would a centralized DB look like
  - Genetics
  - Phenotype
  - Controls?
- How would one transition between existing data and New Resource?
- What steps can be taken to reduce the burden of obviously unfounded claims in the literature?
- How to handle “threads of evidence”

Genetic background

- Background genome can impact significance of the variant
  - Genome “finite” in terms of main effects
  - May as well be infinite in terms of interactions
- Knowledge that a variant is implicated as causative at least once in a genomic context is an important
- Create a meta-database to connect existing elements in an organized and relational way
- Wiki-like environment for independent community curation accompanied by objective and structured database