Measuring the molecular functional consequences of very large numbers of human genetic variants

Jay Shendure, MD, PhD
Department of Genome Sciences
University of Washington

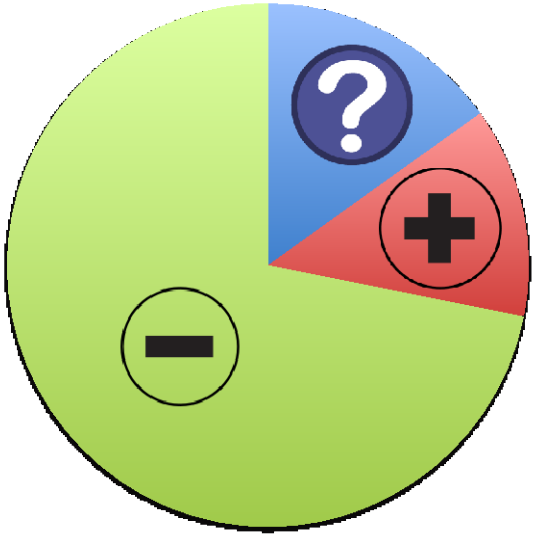
Genome Sciences
UNIVERSITY OF WASHINGTON

The interpretation of genetic variation is a NHGRI mission-critical challenge

Variants of uncertain significance

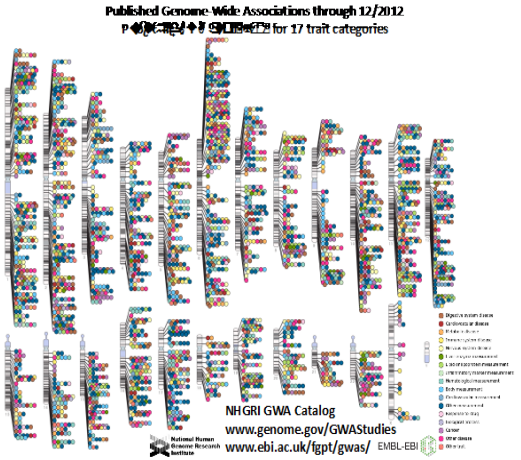
BRCA1&2 → 1M patients

Problem not going away



>1000 genome-wide associations

Small minority for which functional variant(s) known



Computational Prediction?

An Atlas of the Molecular Functional Consequences of Genetic Variation

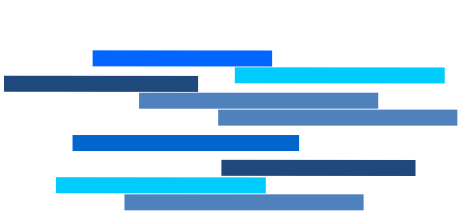
Can we **experimentally** measure the functional consequences of **10 million** human genetic variants?

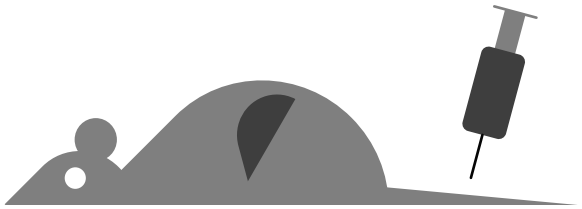
(0.1% of all possible SNVs)

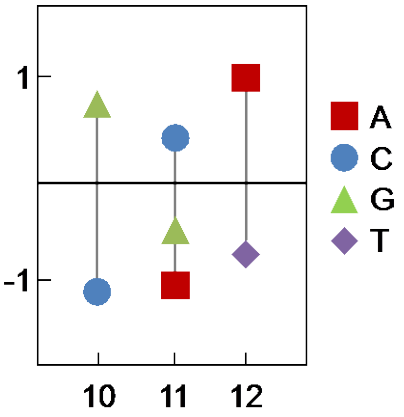
What would we gain?

- **A massive training dataset** for improving algorithms for predicting variant effects
- **Sequence-structure-function maps** for diverse elements (regulatory, protein)
- **The measurements themselves**, for better interpreting Variants of Uncertain Significance and fine-mapping associations

All possible SNVs of an **enhancer**







Doped synthesis & cloning of allelic series of enhancer

↓

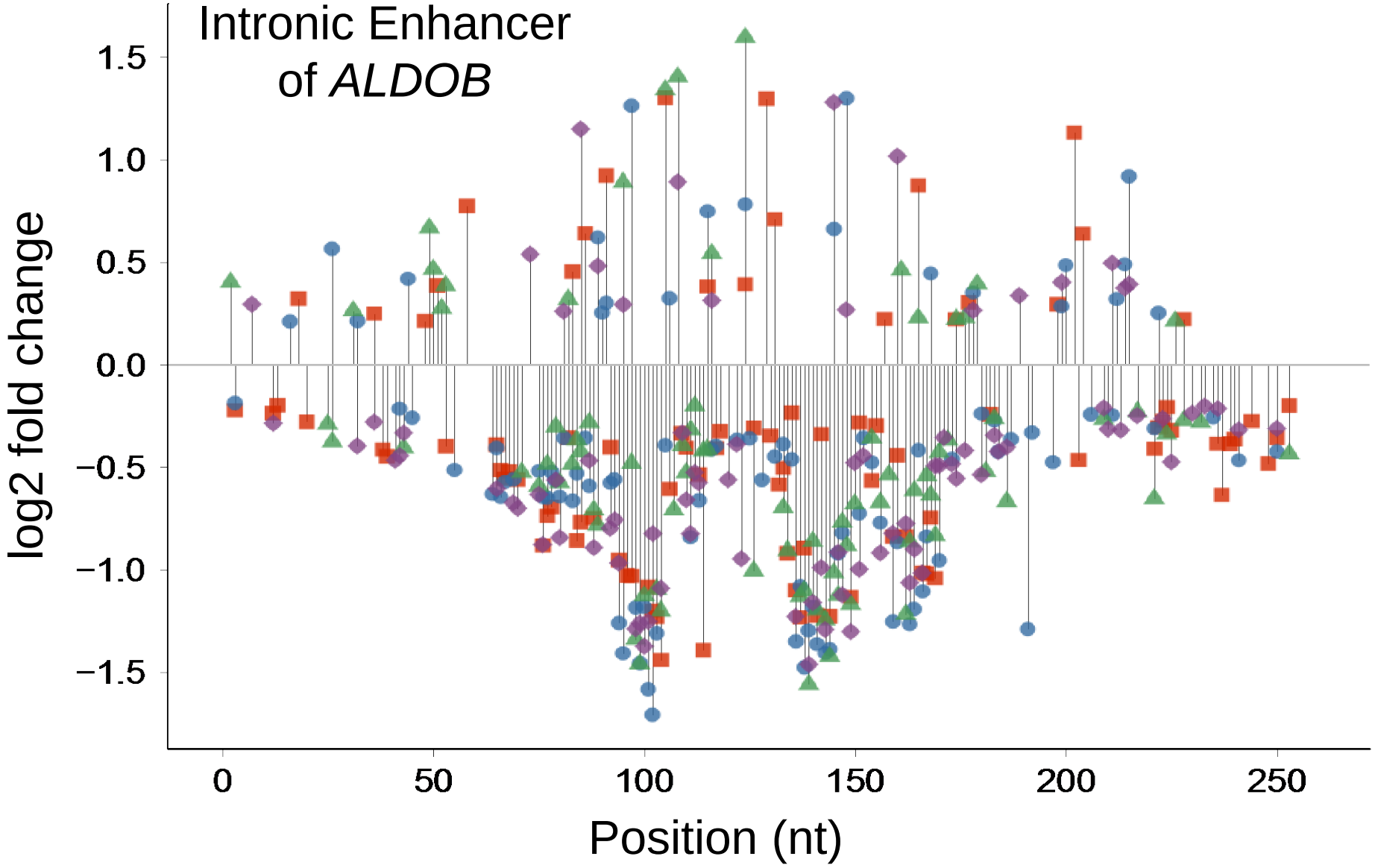
Massively parallel functional assay

↓

Sequence transcribed barcodes & estimate SNV or indel effect sizes

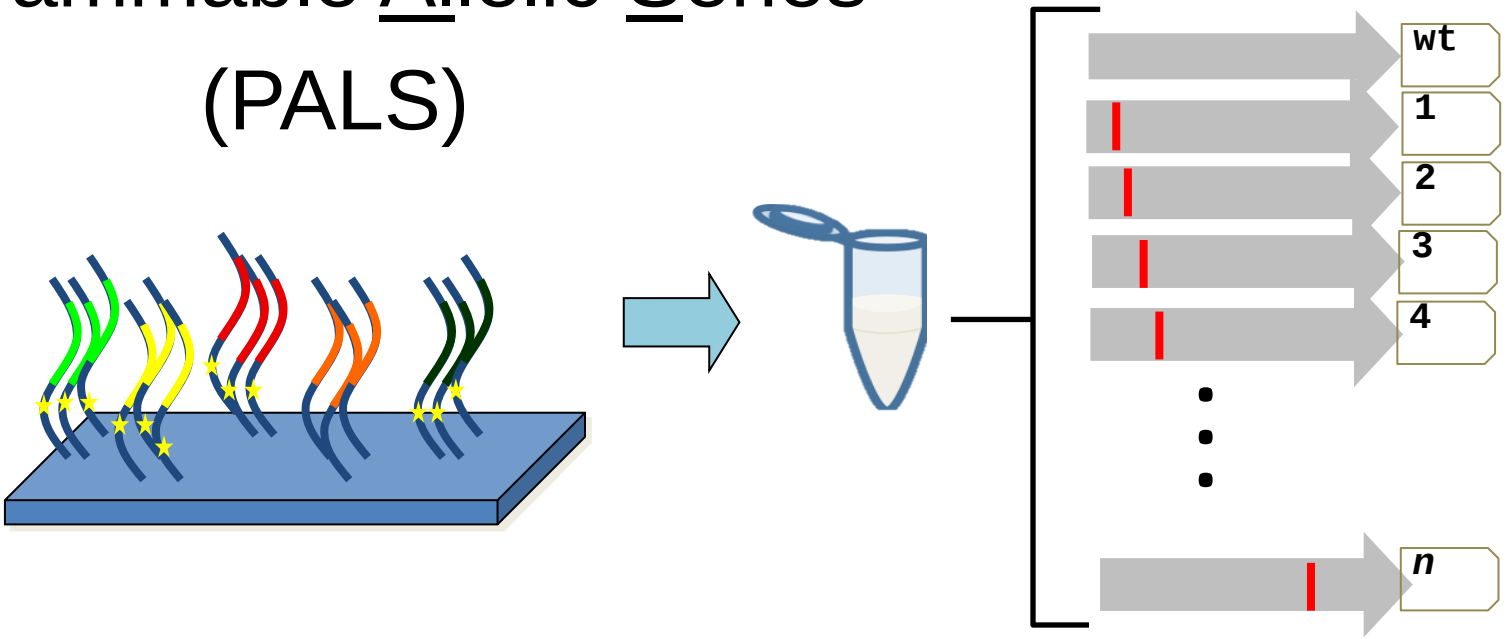
Patwardhan, Hiatt *et al.* Nature Biotechnology (2012)

All possible SNVs of an **enhancer**



All possible codon swaps of a **protein**

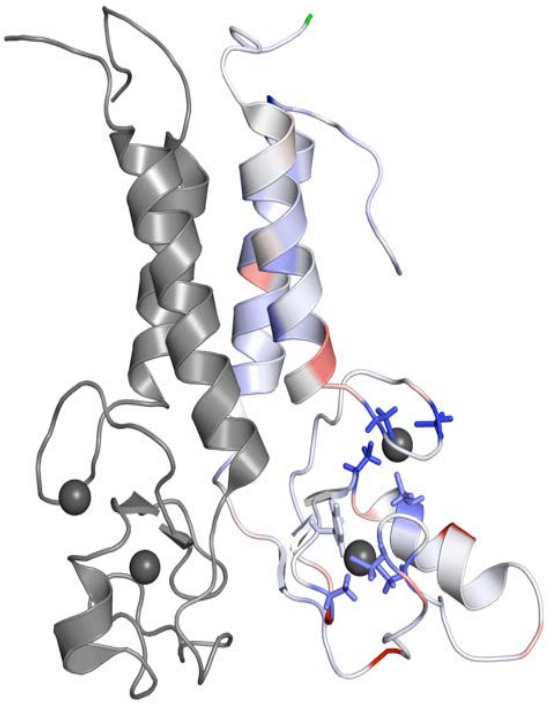
Programmable Allelic Series (PALS)

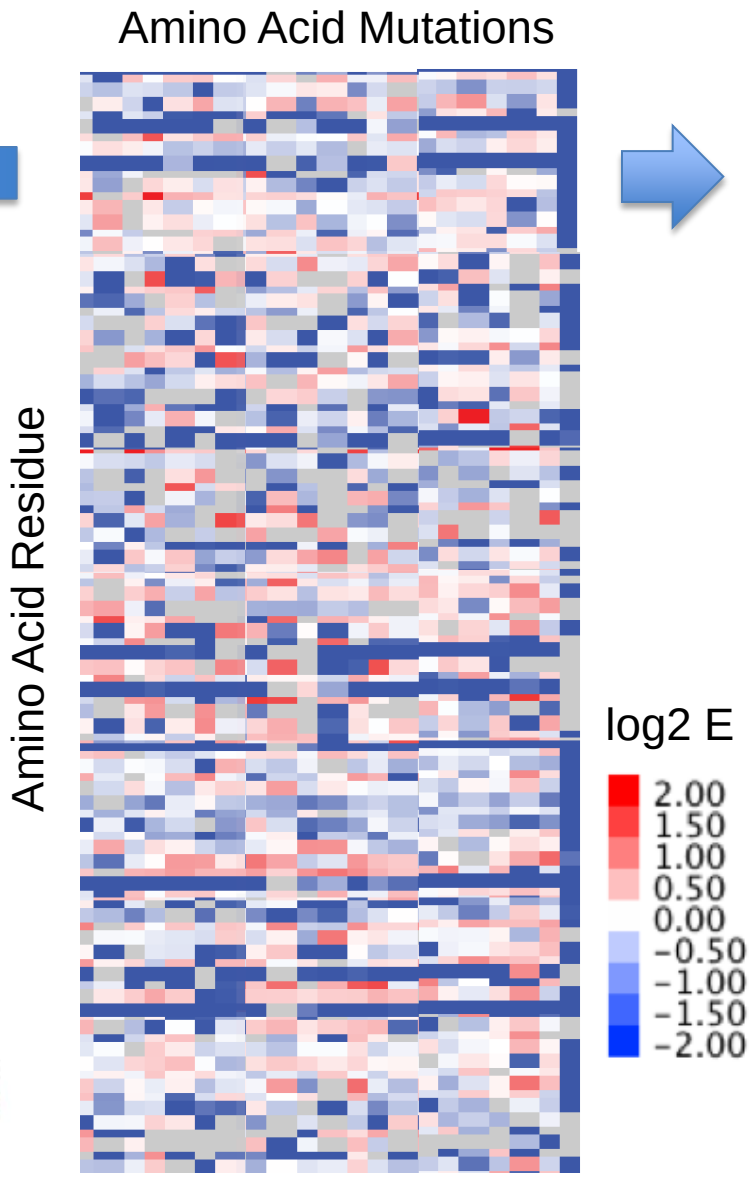


All possible codon substitutions of an ORF in a single, multiplex reaction

Jacob Kitzman

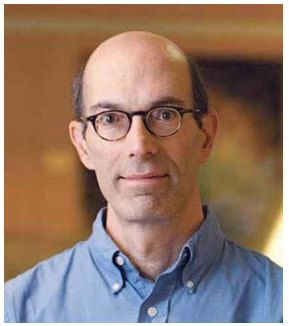
BRCA1 structure-function-pathogenicity





Classification of clinical variants

Lea Starita



Stan Fields

An Atlas of the Molecular Functional Consequences of Genetic Variation

Technologies are rapidly maturing

Rate-limiting factors:

- Allelic Series Production
- Multiplex Functional Assays

Sequencing is the least of the challenges

Which variants?

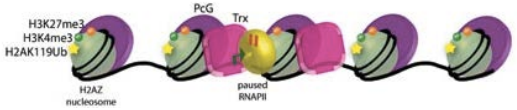
Dense: 2,500 x (1 Kb x 4 mutations) = 1e7

- 500 clinically relevant genes
- 2,000 *cis*-regulatory elements

Sparse: GWAS, eQTL candidates in LD blocks

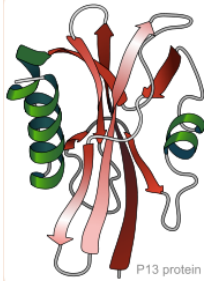
- Better methods are needed
- Much more expensive per variant

Recycle functional assays & allelic series

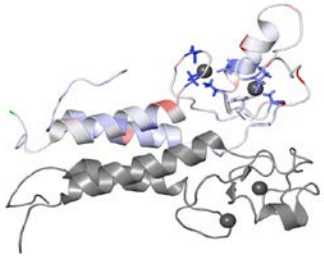


4D x Chromatin

Folding & Stability

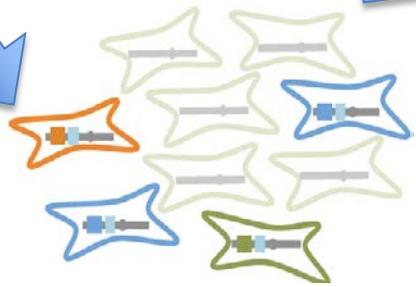


Two Hybrids

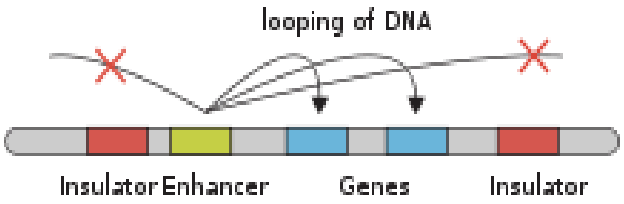


Enhancer Allelic Series

Protein Allelic Series



Cell Type 1
Cell Type 2
Cell Type 3



Transcriptional Activation

Doug Fowler

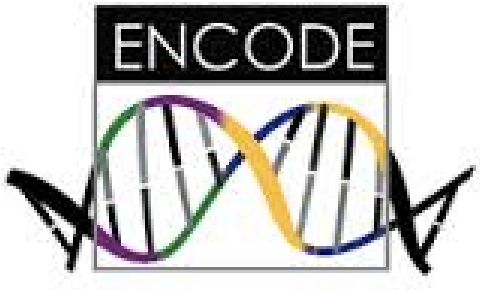
Limitations of Functional Assays

"Context problem" (Botstein & Shortle, Science 1985)

Functional assays ≠ humans

Molecular functionality ≠ pathogenicity

Choice of sequences & contexts informed by:



How would this resource be used?

Rich database of experimental measurements

- Training data for better algorithms
- “Pre-computed” VUS interpretations?

Understanding biology at single-base resolution

- Protein domains, *cis*-regulatory elements
- Learn rules and extrapolate to variants in other members of a functional class

What will success depend on?

- ($15M x 5 yrs) / 10M variants = $30K per 1 Kbp
- Doped oligonucleotide synthesis relatively cheap
- Scaling construction of allelic series
- Multiplex, scalable, recyclable functional assays

- NHGRI: funding, coordination, scaling
- Phase I (100K) → Phase II (1M) → Phase III (10M)
- Technology & assay development throughout