Genomic and Precision Medicine

Week 3: Next-gen sequencing for solving diagnostic dilemmas

Jeanette McCarthy, PhD
Robert Nussbaum, MD

UCSF
University of California
San Francisco
advancing health worldwide™

Lecture 3

- Module 1: Whole Genome Analysis
- Module 2: Clinical interpretation of variants
- Module 3: Using NGS for diagnostic dilemmas
- Module 4: Practical issues

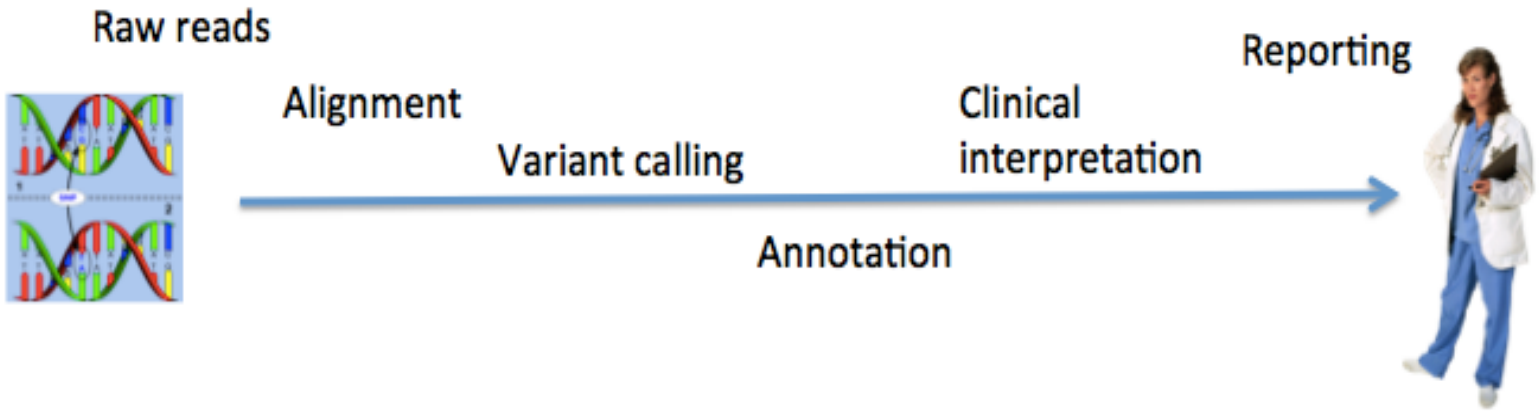
Module 1:Whole Genome Analysis

Whole Genome Analysis

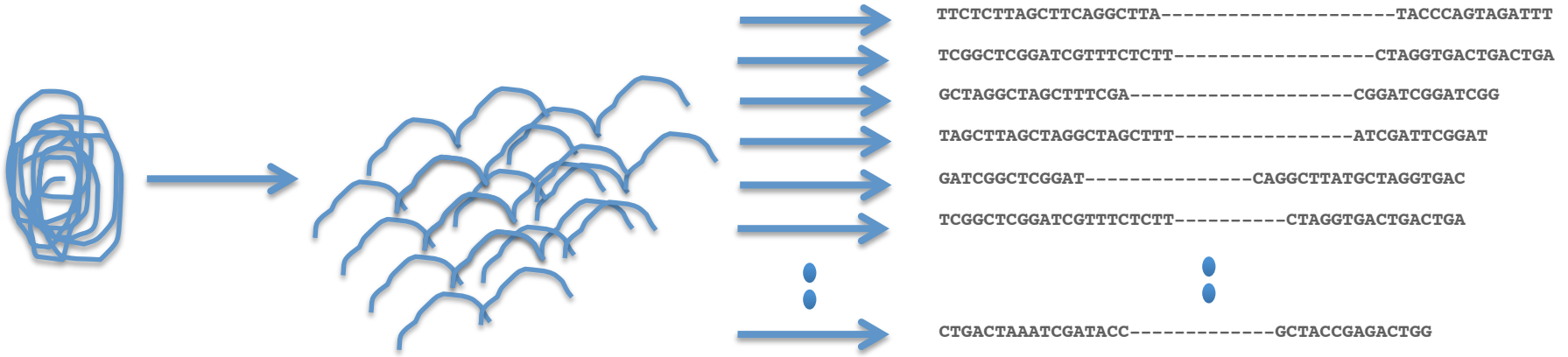
- A genome-wide search for disease-causing variants

- Karyotype – Chromosomes under the microscope
- Cytogenomic Arrays for large deletions/ duplications
- Whole Exome Sequencing (WES)
- Whole Genome Sequencing (WGS)

Going From Sequence to Clinical Use



Sequencing

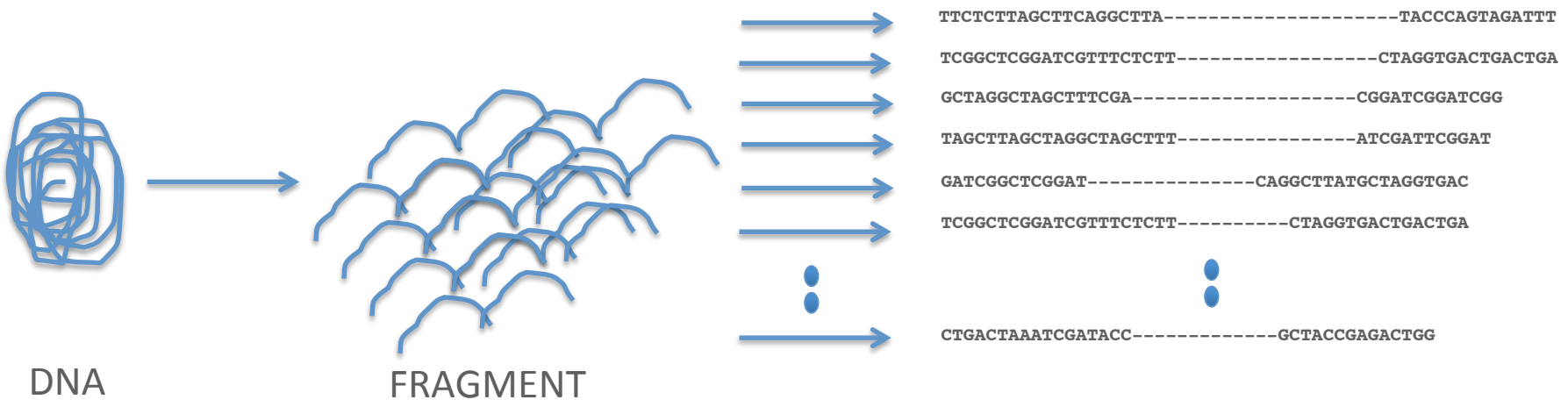


ALIGN ALL SEQUENCES TO THE OTHERS

"De Novo" Alignment

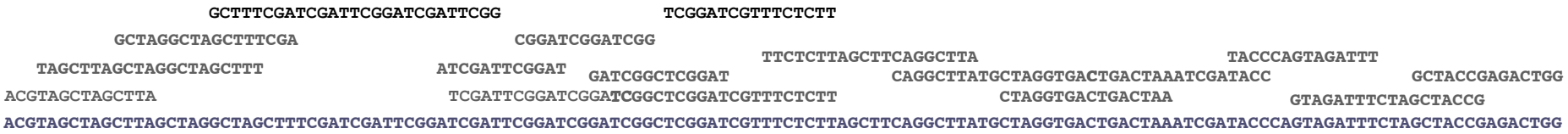
UCSF

Resequencing



SEQUENCE EACH FRAGMENT END

ALIGN EACH SEQUENCE TO THE REFERENCE

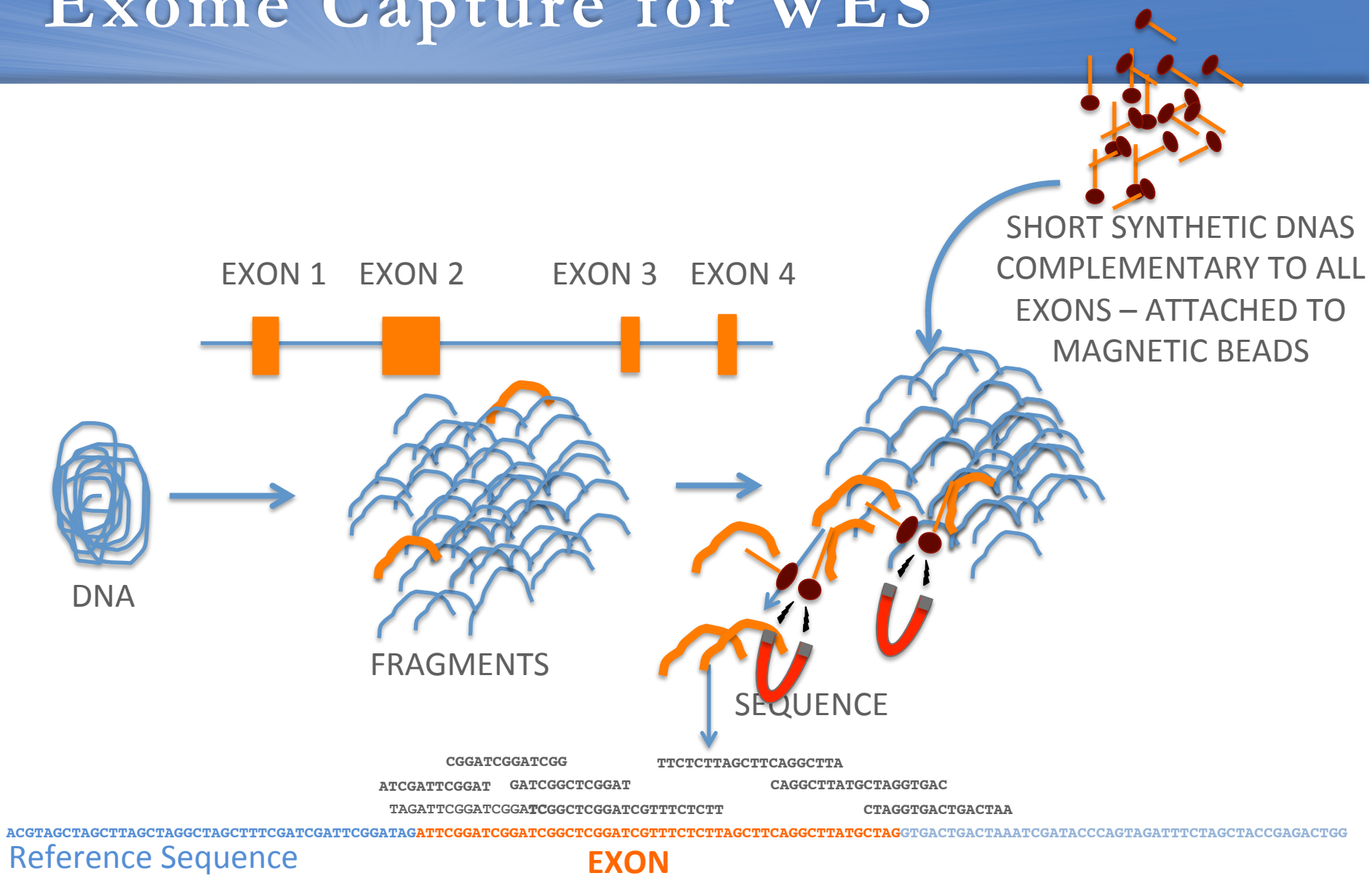


Reference Sequence

There is No Single "Human Genome"

- There is no "normal" or "control" human genome, there are billions of different genomes
- To provide some sort of standard, a "reference genome" was constructed as a consensus among multiple sequences
- Any person's genome differs from the reference at millions of sites, ranging from single nucleotide differences up to hundreds of thousands, even millions of base pairs
- Reference still has gaps in regions where no sequence could be obtained

Exome Capture for WES



UCSF

Whole genome vs whole exome sequencing

Why study just the exome?

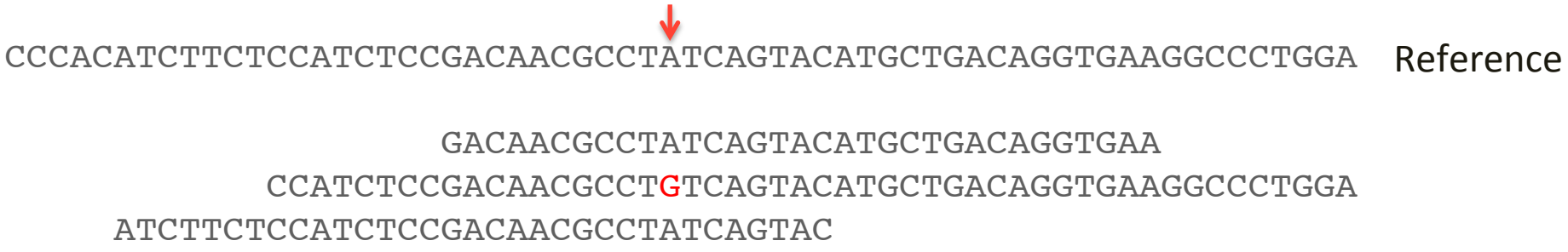
- More predictable effect of mutations
- >85% of known mutations for rare Mendelian disorders occur in the exome
- Cheaper, faster and easier to analyze just 2% rather than the entire genome

What WES can reliably detect

- Small variants (SNVs or small indels) - Read Depth!
- Some CNVs
- Not larger indels or trinucleotide repeats
- Exon deletions are hit-or-miss using depth of coverage measures

WARNING: The technology is evolving rapidly and new advances will change this current snapshot

Read depth (coverage)

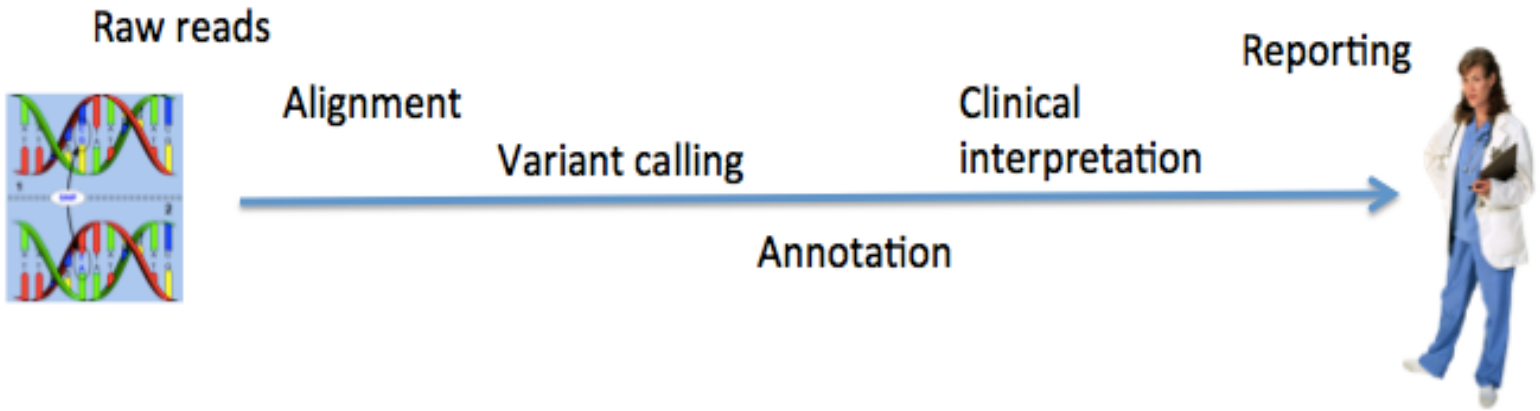


- Would you believe this person is heterozygous (A/G) for a variant with a read depth of 3?

How about Now?

Module 2: Clinical interpretation of variants

Going From Sequence to Clinical Use



Typical Individual Differences from Reference

- ~5-10 million SNVs (varies by population)
  - 40-100,000 SNVs in coding exons
    - 10,000-12,000 synonymous (no amino acid change)
    - 8,000-11,000 non-synonymous, in 4,000-5,000 genes
- 200,000-300,000 indels 1-20 bp (varies by population)
  - ~150 in-frame indels in exons
  - ~200-250 shift the reading frame of an exon
- 500-1000 CNVs >1,000 bp

Approximate

Basic annotation of variants

- Gene name (if in a gene)
- Chromosome location of the change (position in reference genome)
- Location of the change within the mRNA/cDNA
- Location of the amino acid change in the protein
- Effect on protein (if in a gene)

Gene	Chr.	Genomic	cDNA	Protein	Effect
BRCA1	17	g.37038192G>T	c.199G>T	p.Gly67Trp	Non-synonymous
BRCA1	17	g.37042469_37042470delTG	c.231_232delTG	p.Cys77Ter	Stop-gained

Advanced annotation of variants

- Variant dependent methods
  - Allele frequency
  - Predicted effect of variant on protein
  - Evolutionary conservation, protein structure, amino acid properties
  - Functional characterization of variant (in vitro and/or in vivo)
- Disease-dependent methods
  - Mode of inheritance
  - Cosegregation with disease is families
  - Prior association of the gene with disease
  - Pathway analysis

Criteria used to evaluate variants

- Variant dependent methods
  - Allele frequency
  - Predicted effect of variant on protein
  - Evolutionary conservation, protein structure, amino acid properties
  - Functional characterization of variant (in vitro and/or in vivo)
- Disease-dependent methods
  - Mode of inheritance
  - Cosegregation with disease is families
  - Prior association of the gene with disease
  - Pathway analysis

Allele frequency in general population

- For a suspected Mendelian disease, a variant observed in the general, healthy population is assumed non-pathogenic
- 1000 Genomes (www.1000genomes.org)
- dbSNP
- And others…..

Predicting the effect of a variant is CHALLENGING

Probably Damaging

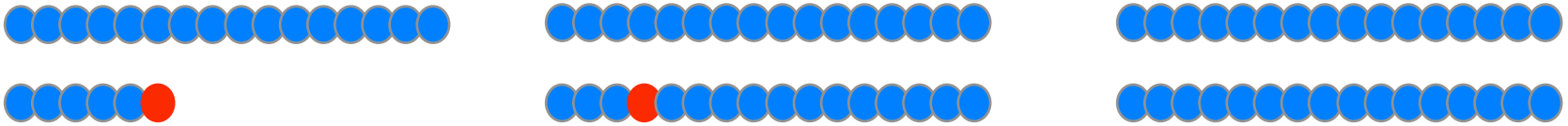
- Stop-loss
- Stop-gained
- Frameshift
- Splice disruptor

Possibly Damaging

- Non-synonymous
- In-frame In/Del

Likely not Damaging

- 5’/3’ UTR
- Synonymous
- Intergenic
- Intronic
- Non-coding genes

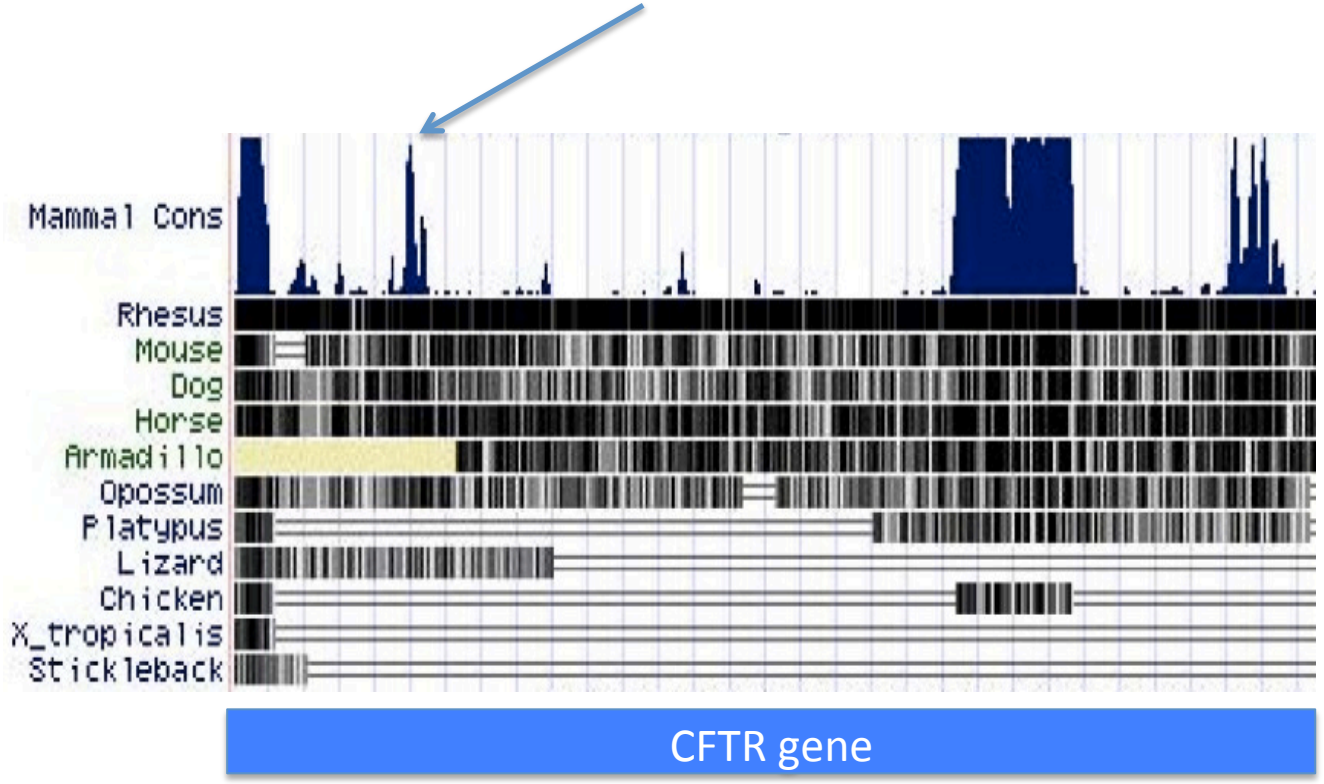


Predicting the effect of non-synonymous variants

- Evolutionary conservation
- Protein structure
- Amino acid properties
- These criteria are applied together by various computer algorithms to assess how damaging a change might be

Evolutionary conservation

Mutations in conserved positions more likely deleterious

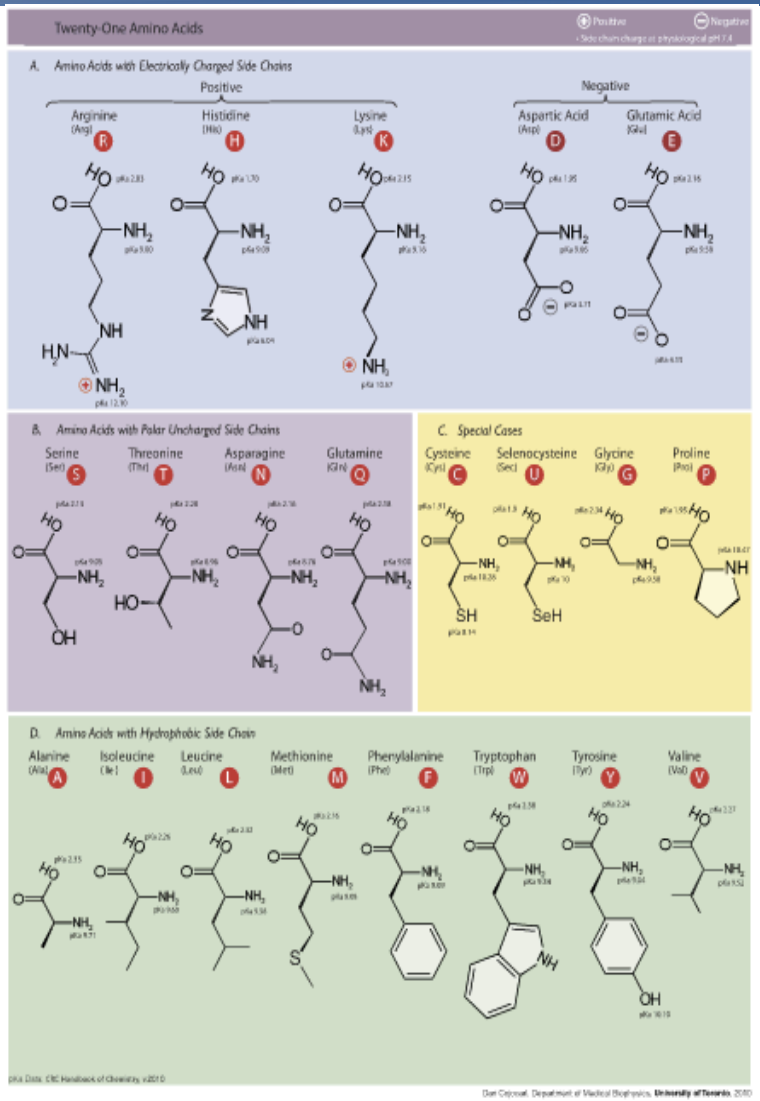


Amino acid properties and protein structure

Not all amino acid changes are equivalent

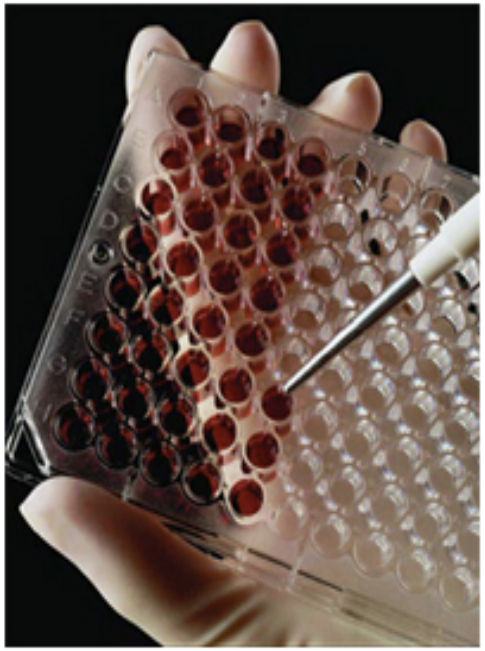
Conservative changes less likely to affect protein structure/ function

Location of change within protein matters as well



In vitro/in vivo functional studies

In vitro (cell-based assays)



In vivo (animal models)

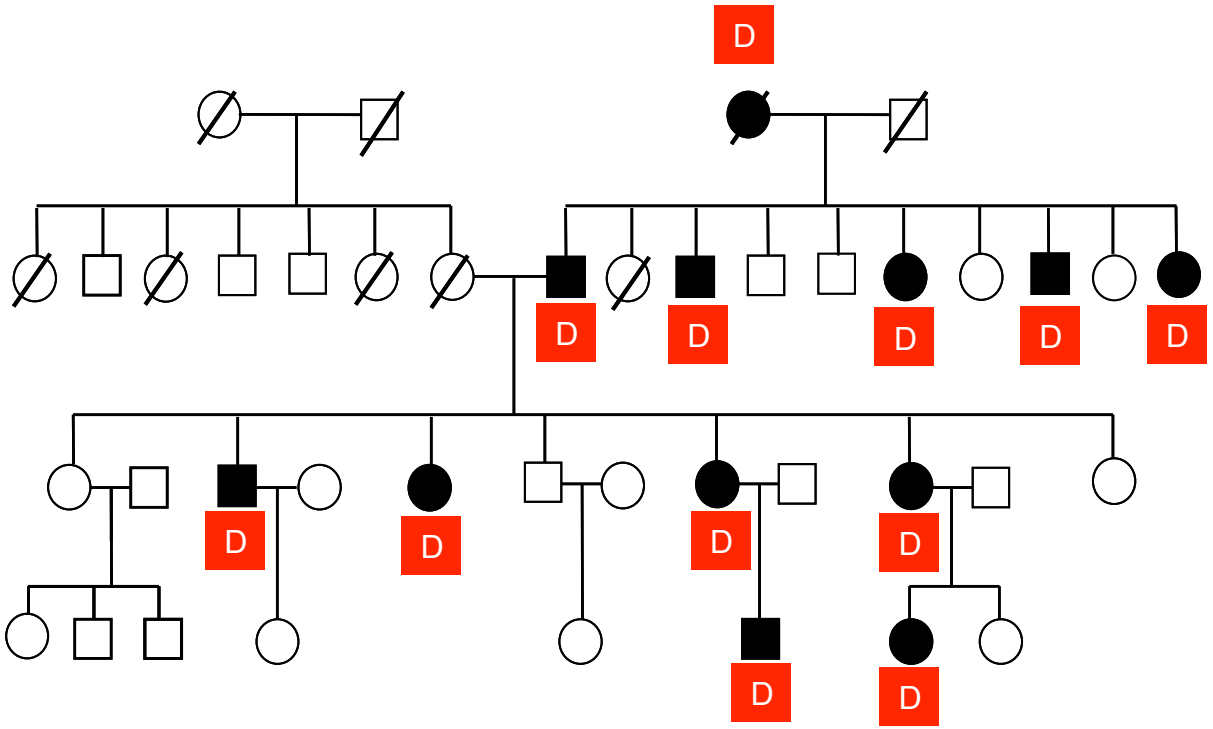
Criteria used to evaluate variants

- Variant dependent methods vary in their ability to predict the effect of a variant on gene or protein function. Some are highly predictive, others are, at best, suggestive or circumstantial.

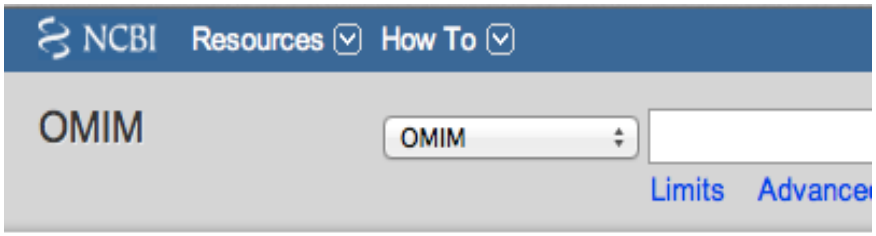
Criteria used to evaluate variants

- Variant dependent methods
  - Allele frequency
  - Predicted effect of variant on protein
  - Evolutionary conservation, protein structure, amino acid properties
  - Functional characterization of variant (in vitro and/or in vivo)
- Disease-dependent methods
  - Cosegregation with disease is families
  - Prior association of the gene with disease (OMIM)
  - Pathway analysis

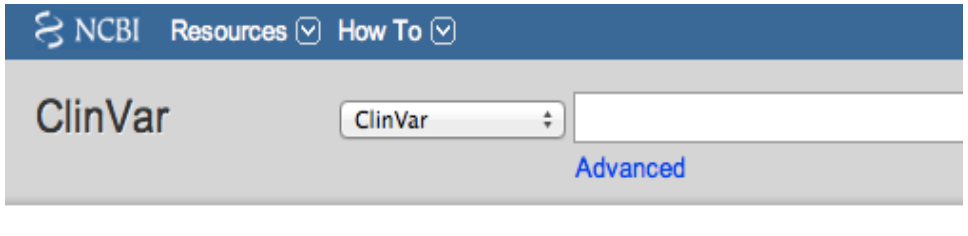
Cosegregation of variant with disease in families



Prior association of gene with disease



http://www.ncbi.nlm.nih.gov/omim



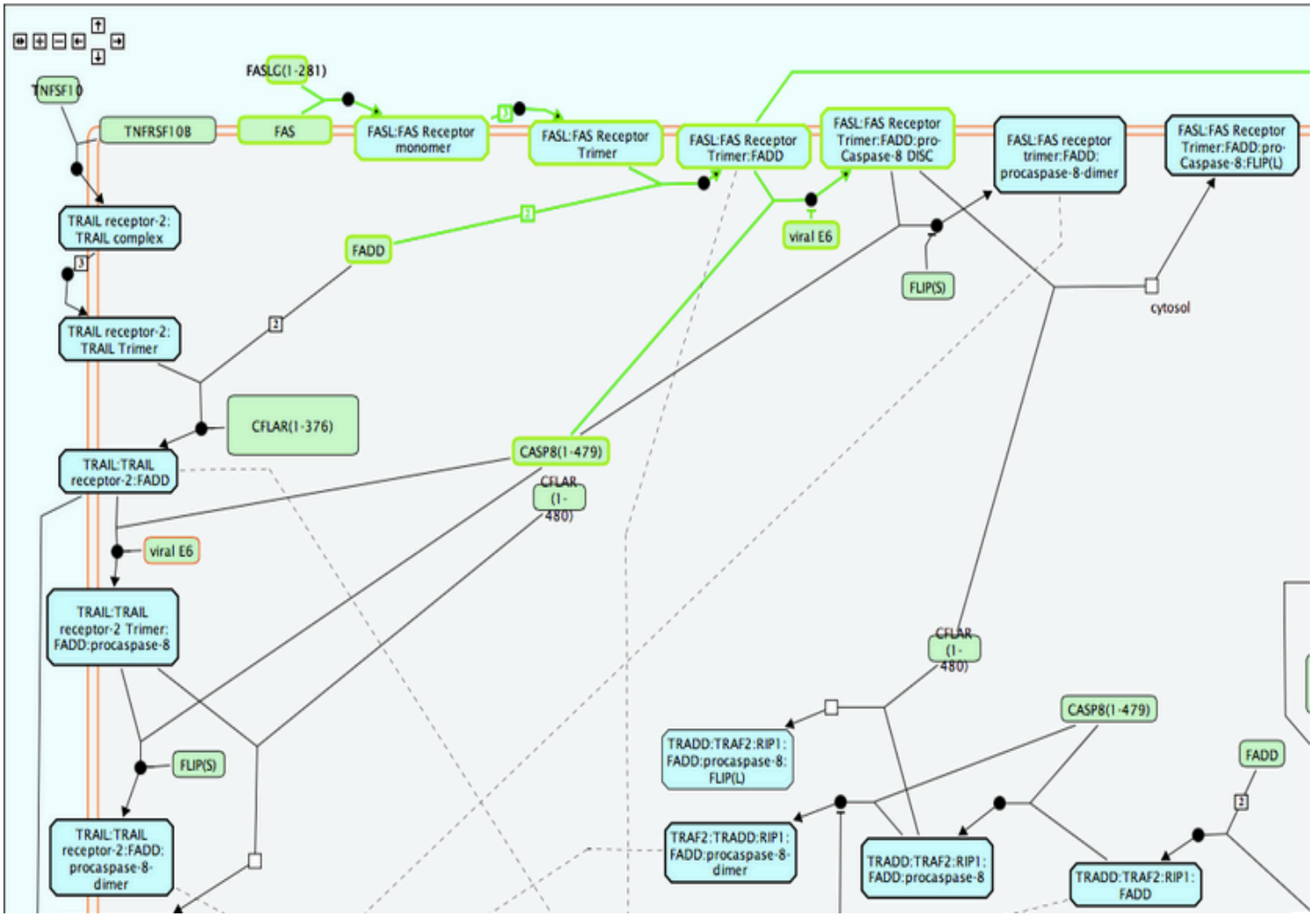
https://www.ncbi.nlm.nih.gov/clinvar/

HGVS
HUMAN GENOME VARIATION SOCIETY

Locus Specific Mutation Databases

http://www.hgvs.org/dblist/dblist.html

Gene in a disease pathway



Typical classification scheme

- Known pathogenic
- Likely pathogenic
- Variant of unknown significance (VOUS or VUS)
- Likely benign
- Benign

Classifying Variants

Deleterious Variant	Pathogenic Variant
Located in or near a coding exon and is….. • A frameshift or nonsense mutation that causes premature termination of translation	• Strongly associated with disease in affected individuals versus in unaffected individuals
• A non-synonymous amino acid change affecting a highly conserved residue through evolution	• Tracks with the disease in a family with multiple affected members
• A splice-site mutation in an intron that is highly likely to cause abnormal splicing	• Experimental evidence in animal models that the alteration causes disease
• NOT found as a common variant in a population of matching ethnic background	• *In vitro* experiments showing the variant changes function and is likely to cause disease

Classifying Variants

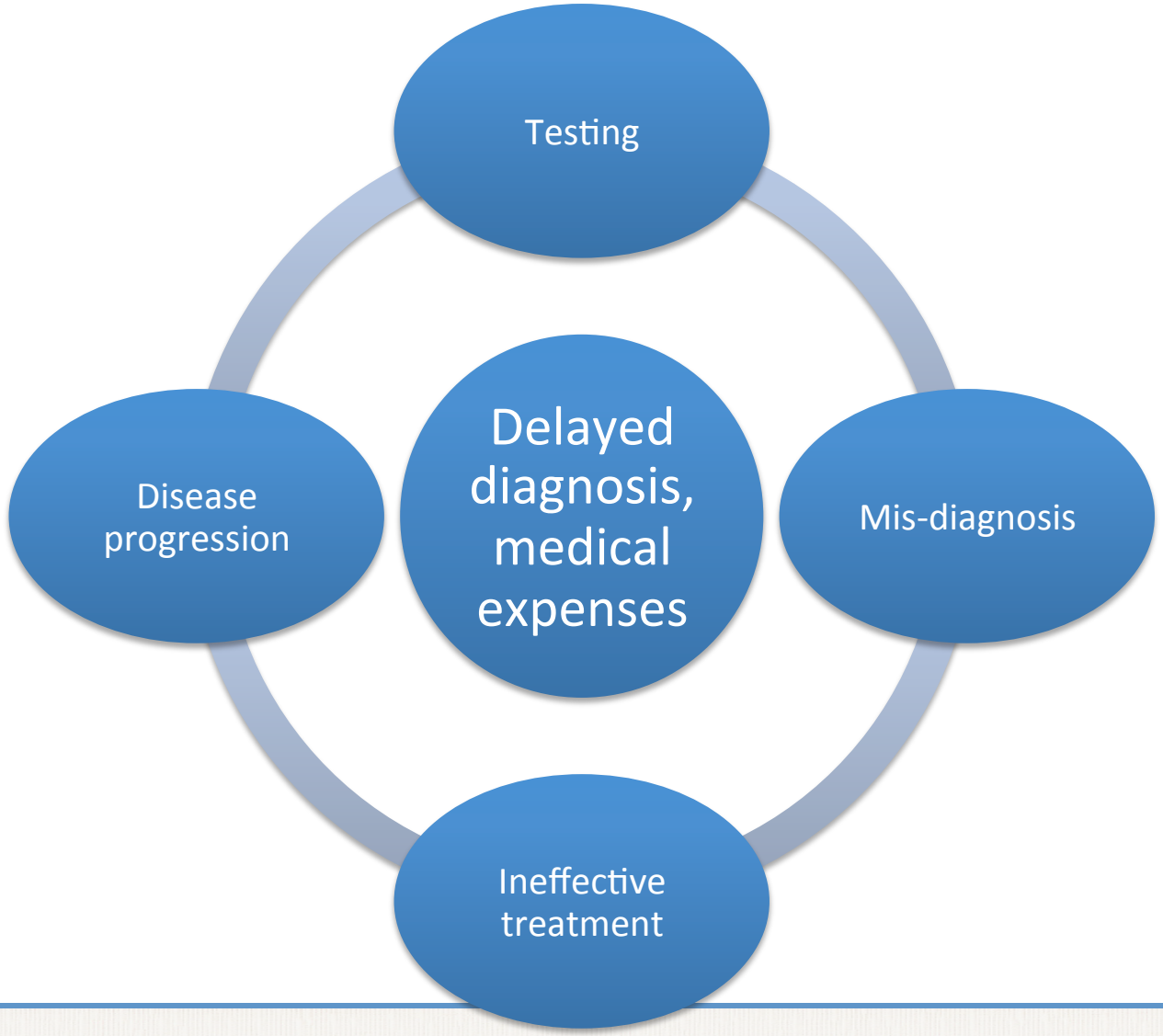
(Likely) Benign Variants
Common in an ethnic group without being associated with frequent disease
Synonymous change that does not change the amino acid encoded
Non-synonymous change of an amino acid residue that is not conserved between species

"Typical" Variant Classification Test Report

- "Mutations in this gene have been previously reported to cause this disease"
- "This variant affects a highly conserved cysteine residue in gene XYZ resulting in the substitution of a positively charged amino acid, arginine, for a sulfur-containing amino acid in an important functional domain of the enzyme encoded by gene XYZ"
- "Published biochemical studies have shown this substitution causes loss of activity of the enzyme"
- "This variant has been reported in 3 unrelated patients with this disease"
- "In one unrelated patient, the disease was co-inherited with the mutation in 5 other affected members of the family"

Module 3: Using NGS for diagnostic dilemmas

A typical diagnostic odyssey



Ending a "Diagnostic Odyssey"

Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease

Elizabeth A. Worthey et al. *Genet Med* 2011:13(3):255–262.
Medical College of Wisconsin

Intractable inflammatory bowel disease

- Whole Exome Sequencing of 15 month old male infant
- ~16,000 total variants within coding portion of genes (SNVs, small duplications or deletions. ~1500 were "novel"
- ~7,100 were non-synonymous substitutions, premature stops, or small insertions or deletions in coding exons of which , ~ 1,100 were novel, not present in databases of normal variants
- 136 variants fit an autosomal recessive (AR) or X-linked (XL) inheritance model
- 1 variant among the 136 that fit AR or XL inheritance, altered a conserved amino acid, was predicted to be damaging, was not present in reference genome, and was not in a gene in which deleterious mutations causing a different phenotype was known.

E.A. Worthey et al. Genet Med 2011:13(3):255–262.

Mode of inheritance

- Dominant
- Recessive

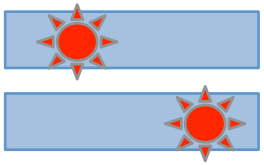
One co… …gene affected

XIAP

Indication for bone marrow transplantation

Heteroz…

Compound heterozygote



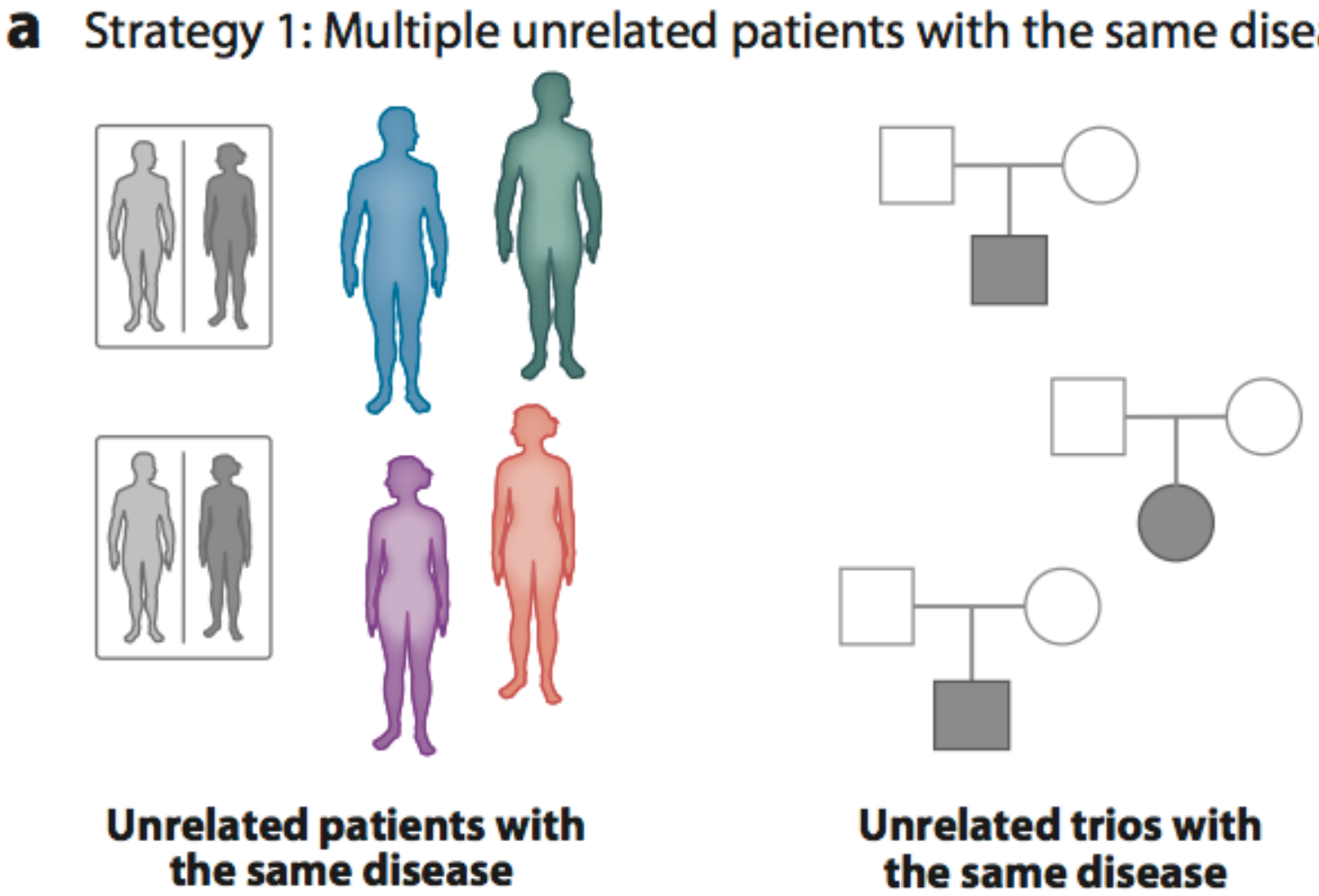
X-linked

~7,000 potentially damaging variants in genes, 1,100 of them novel

136 fit these inheritance patterns

Using unrelated patients

Boycott et al. Ann Rev Med 2014



The Trio in Whole Genome Analysis

Original Article

Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders

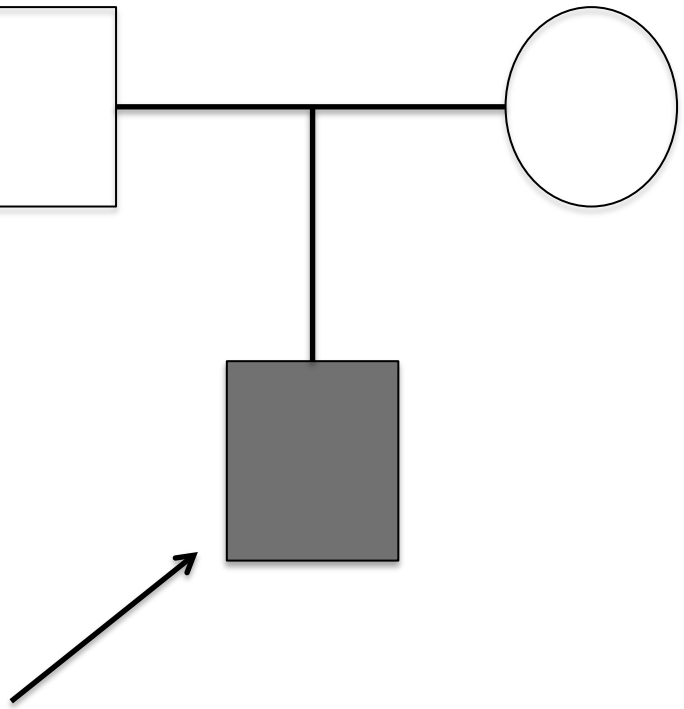
Yaping Yang, Ph.D. et al. .

250 consecutive patients with undiagnosed diseases undergoing trio analysis

N Engl J Med
Volume 369(16):1502-1511
October 17, 2013

UCSF

The Trio in Whole Genome Analysis



Mode of inheritance

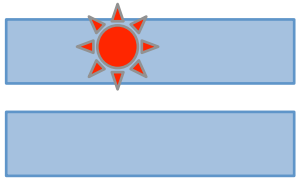
- Dominant

One copy of gene affected

- New Mutation

New variant in the patient not present in either parent

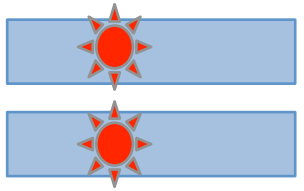
Heterozygote



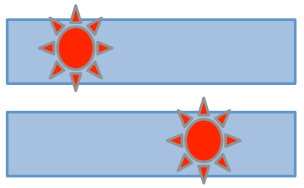
- Recessive

Both copies of gene affected

Homozygote

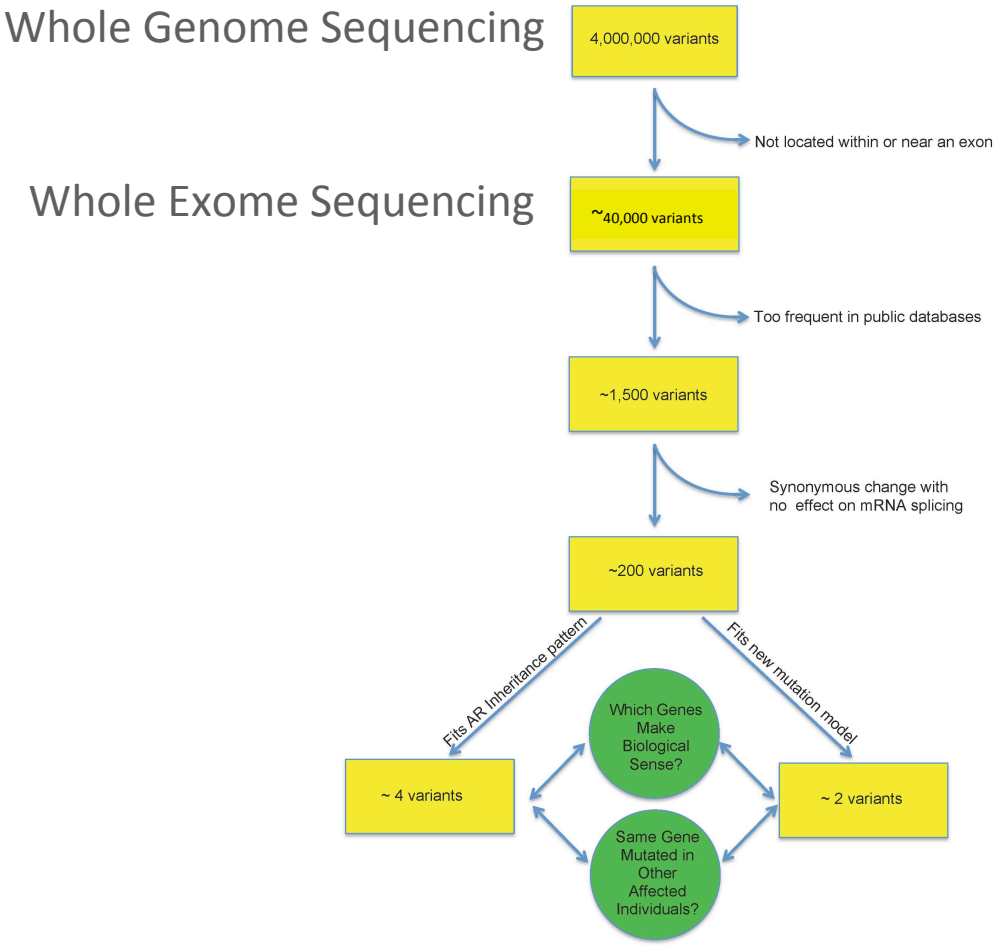


Compound heterozygote



X-linked

Representative Filtering Scheme



Which patients are getting WES

Disease Category	Age at Time of Testing				
	Fetus	< 5 yr	5-18 Yr	>18 yr	Total
Neurological (+/- other disorders)	1	110	86	16	213
Non-Neurological	3	14	8	12	37
Total	4	124	94	28	250

Yang Y et al. N Engl J Med 2013;369:1502-1511

Diagnostic yield in 250 WES patients

- The underlying genetic defect was found in 62 (25%) of 250 consecutive patients
  - 33 were dominant (with 29 new mutations)
  - 16 recessive
  - 9 X-linked
  - 4 had two diseases

Yang Y et al. N Engl J Med 2013;369:1502-1511

WES to identify a gene for MFDM

- Tested 4 unrelated individuals with mandibulofacial dysostosis
- Assumptions: All individuals would have a mutation in the same gene (not necessarily the same mutation). Condition is rare

4th patient had probable deletion in *EFTUD2*

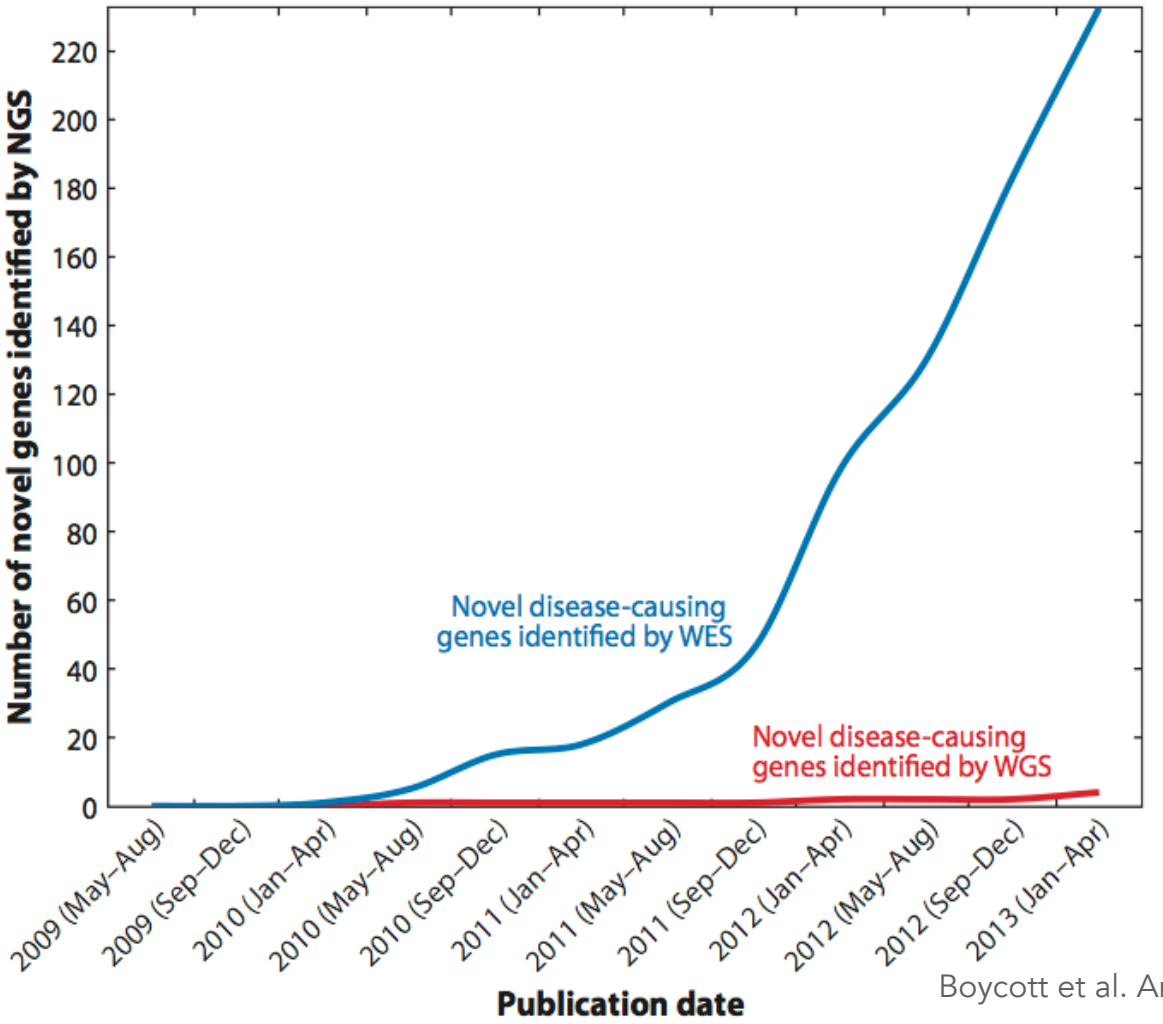
Included *EFTUD2*

Gene with SNV in N patients, N=1,2,3,4	1	2	3	4
Gene mutated in N patients, N=1,2,3,4	1	2	3	4
Missense, nonsense, in/del, splice	2,500	200	20	2
Allele frequency ≤ 1%	1,500	160	8	0

MUC4

Lines et al. Am J. Hum. Genet. 2012

Identification of disease genes using NGS



Boycott et al. Ann Rev Med 2014

Module 4: Practical aspects

The "Incidentalome"

- Unanticipated Pathogenic Variants: Variants that appear deleterious and might be of significance but were not what was originally being looked for.
- Chosen to be "actionable" - would improve the care of the patient
  - Example: Finding a hereditary cancer predisposition gene during the search for the cause of a neurodegenerative disease
- Are these Fortunate Discoveries or False Alarms?
- How much real good did you do uncovering these?
  - Actionability vs. Unnecessary anxiety & higher health care costs

Incidental findings during WGS

American College of Medical Genetics and Genomics recommendations (revised 2014)

- Unless the patient specifies otherwise, the laboratory must report any clearly pathogenic mutations in one of an initial set of 56 genes, regardless of the age of patient or the indication for which testing was originally ordered.

Where to get tested

- NIH Undiagnosed Disease Program (UDP)
- Rare Genomics Institute
- Academic medical center laboratories
- Many commercial options

Informed consent – proposed minimum elements to include

- Scope and Description – What is being tested
- Benefits
- Risks
- Testing is Voluntary
- Alternative test
- Confidentiality
- Future use
- Incidental findings

Ayuso et al., European Journal of Human Genetics (2013) **21,** 1054–59

Informed consent for WGA – What must patients be told?

1. As with all medical care, the test is voluntary
2. What other testing options are available?
3. We may not find the cause of your condition
4. The study has limitations – we cannot find everything that may be significant to you
5. The study will uncover changes in your genes that we are not able to interpret today
6. If uninterpretable variants are later shown to be disease-causing, we cannot guarantee we can find you to give a revised report

Informed consent for WGA – What must patients be told (Part 2)?

7. The study may uncover changes in your genes that we think are of medical/clinical importance but are not the reason we sent the test – do you want them?
8. We may uncover variants that are of clinical significance not only to you but to other family members
9. Information will be saved and protected as with all other personal health information but it will be used for quality improvement and quality assurance purposes
10. There is a risk of unintended disclosure and, where applicable, impact on insurance or employment not currently protected under non-discrimination laws
11. There is a risk we may uncover variants that indicate family relationships are not what they are currently understood to be

The Bottom Line

- Whole Genome Analysis is complex at many levels including
  - Technology – limitations
  - Interpretation
  - Patient consent