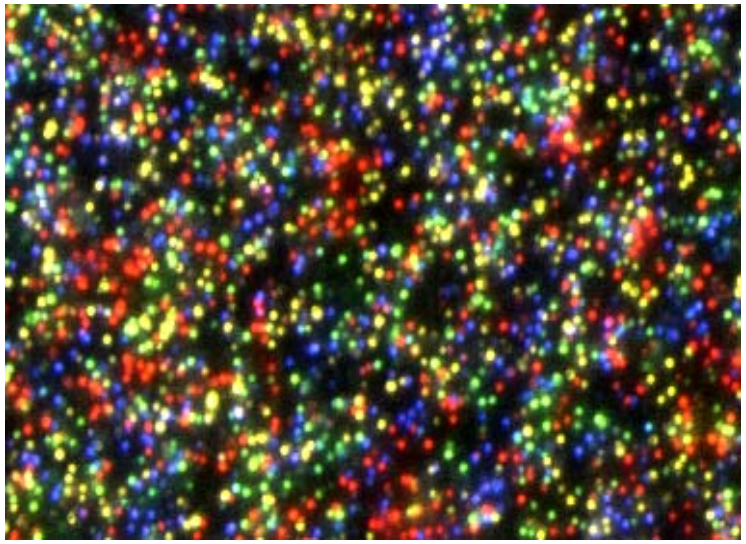
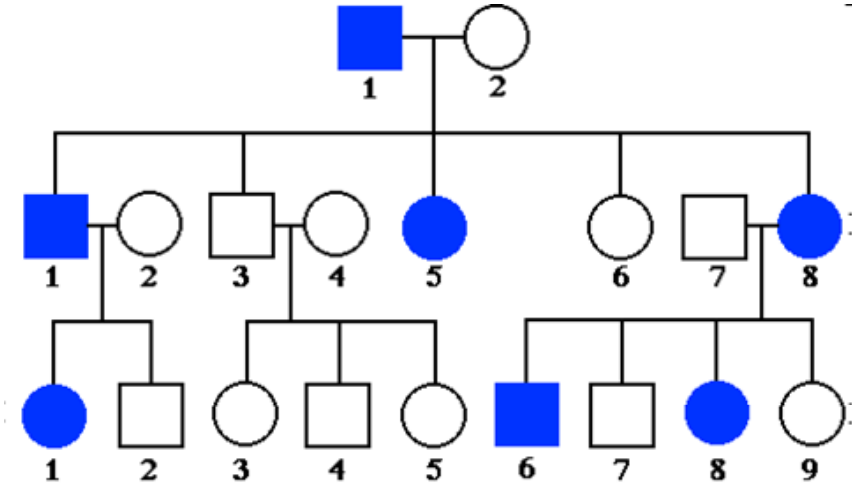
Next Generation Mendelian Genetics by Exome Sequencing

Jay Shendure, MD, PhD
Dept. of Genome Sciences
University of Washington

Second generation sequencing

- 10,000-fold drop in the cost of DNA sequencing since 2005 (& continuing to fall)
- How are these technologies best applied to key problems in human genetics?





Exome = Protein Coding Genome

- split amongst >160,000 exons
- total length of **~30 megabases**
- **include** canonical splice-sites, miRNAs
- **exclude** 5’ & 3’ UTRs (untranslated)

Exome Resequencing

~1% of human genome ~20x cheaper	No structural variation! No non-coding variation!

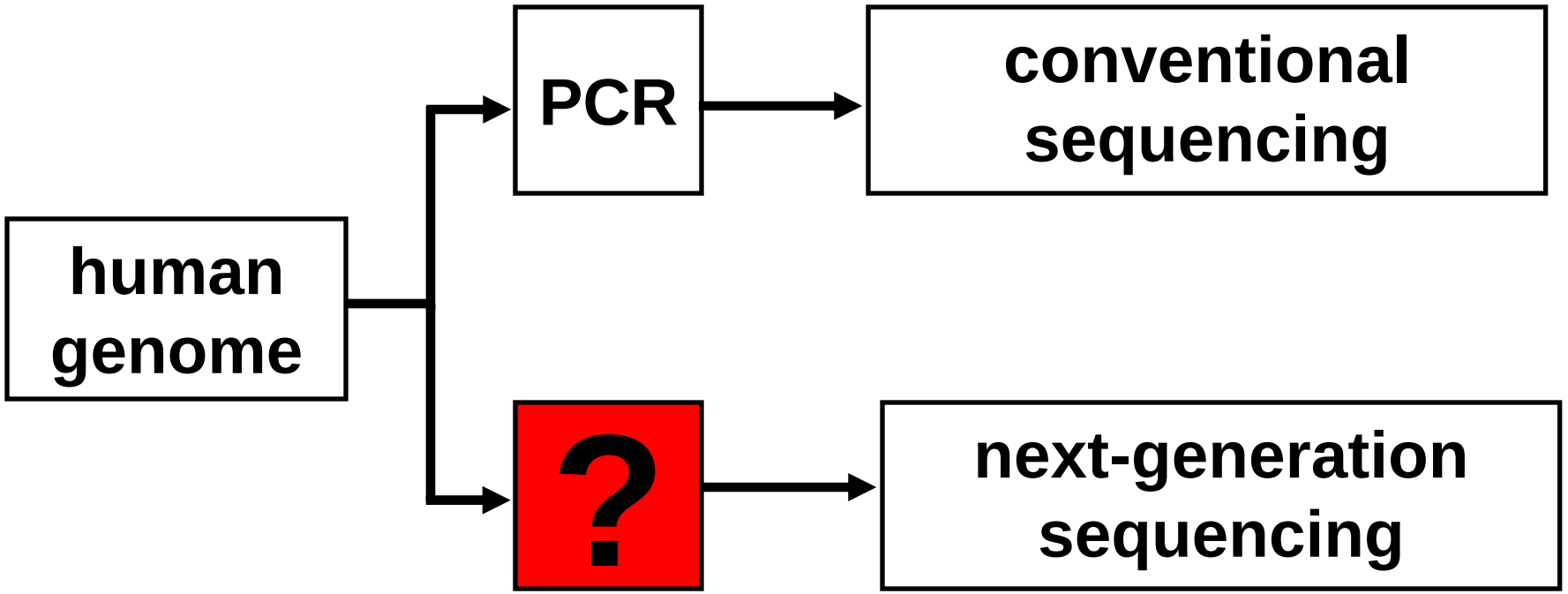
nonsynonymous variation

functional validation

implicated = causal

Technical challenge

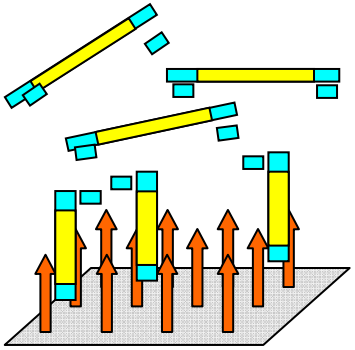
one exon



>160,000 exons

Many ways to skin a cat…

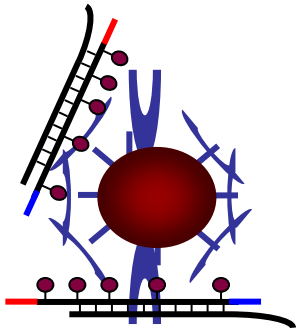
- Array hybridization



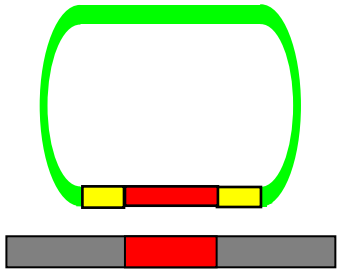
Agilent Technologies

NimbleGen SYSTEMS, INC.

- In solution hybridization

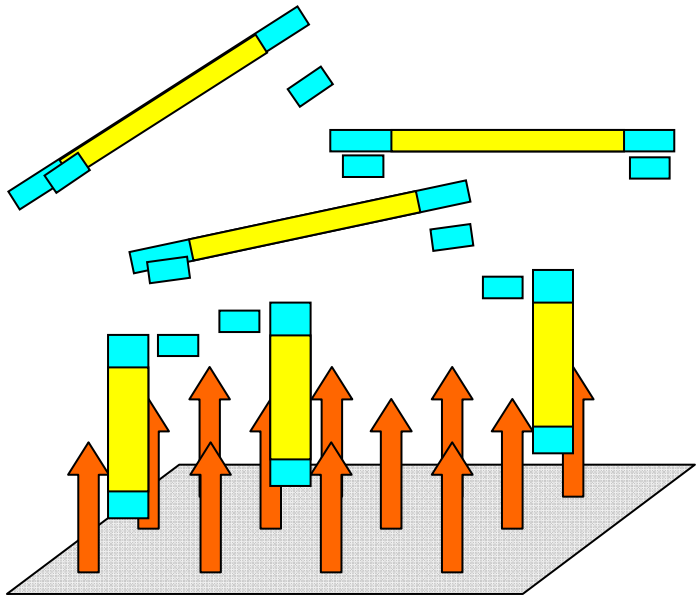


- Molecular inversion probes

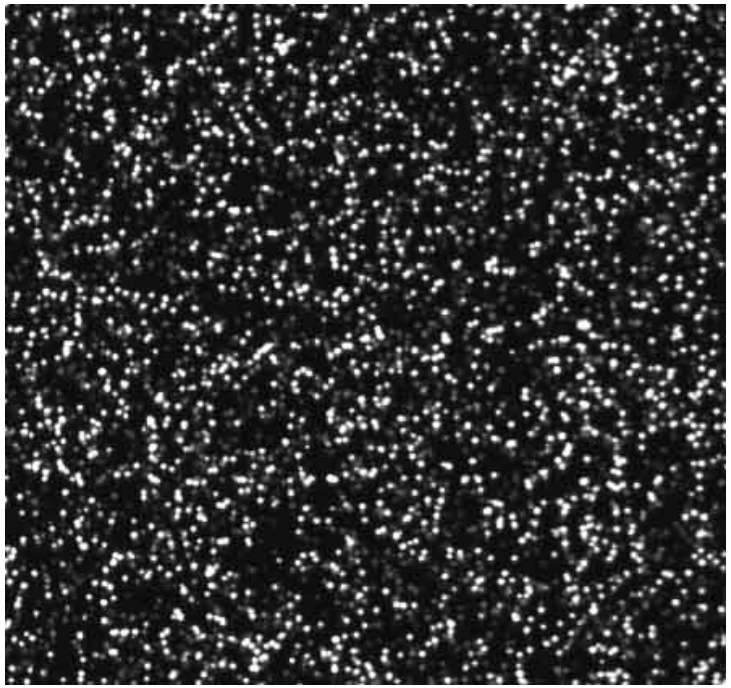


Exome capture by hybridization

shotgun genomic library

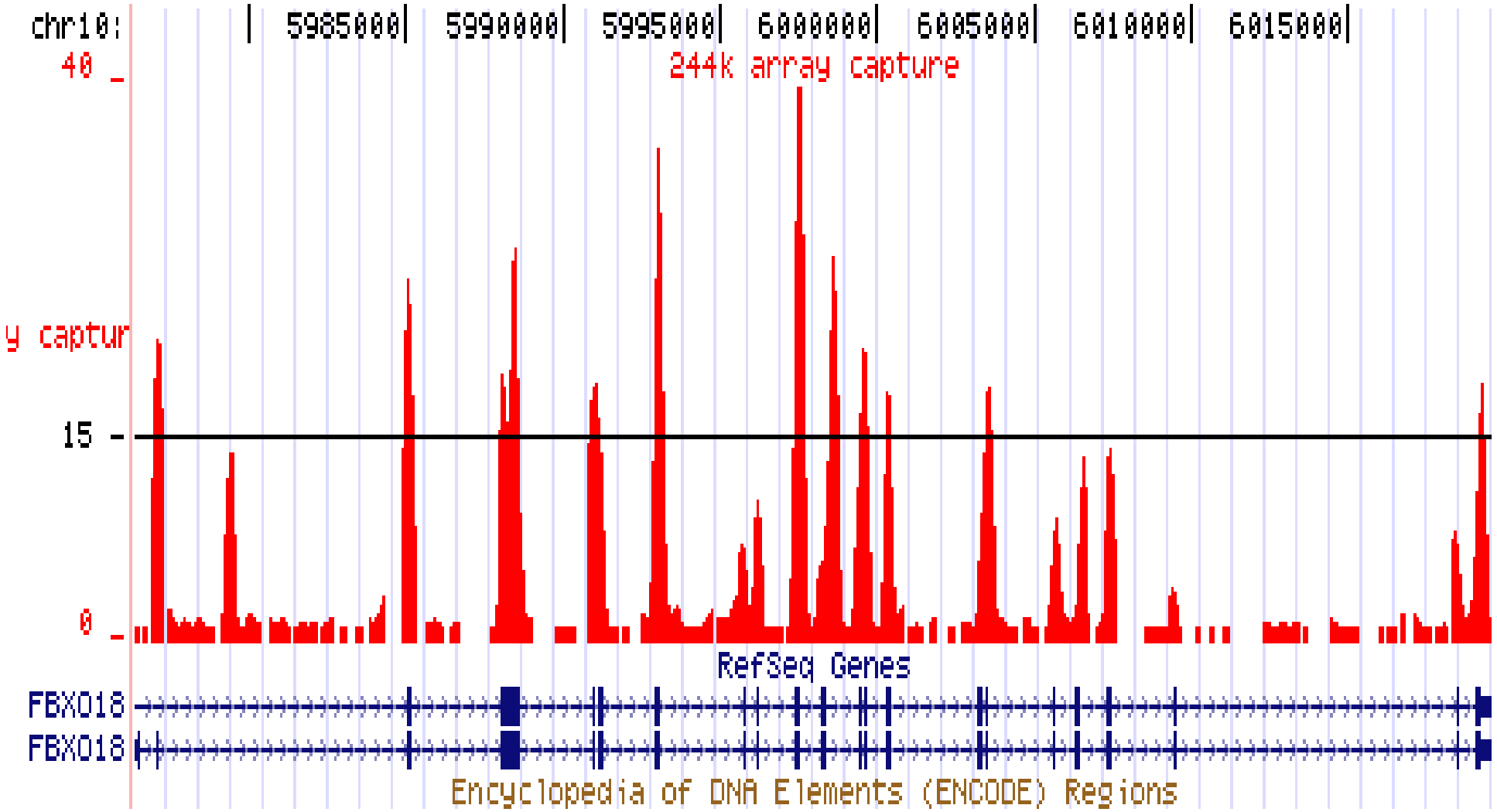


exome microarray



massively parallel sequencing of enriched library

Enrichment for reads overlapping exons



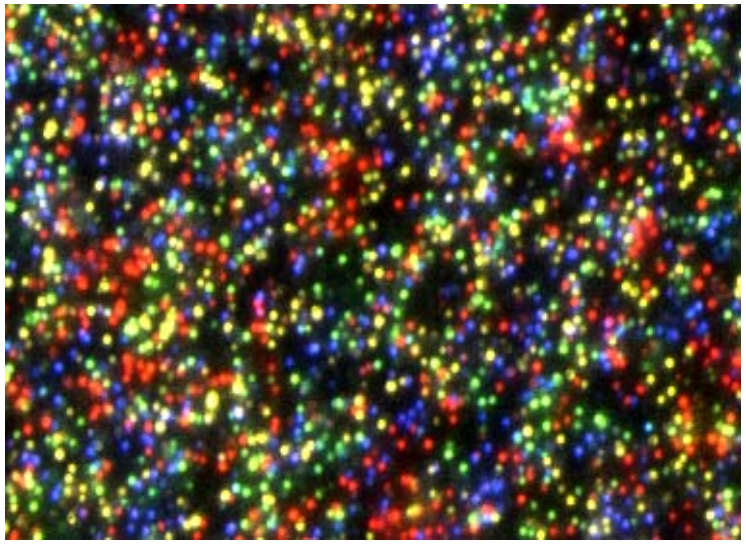
Consistency by population

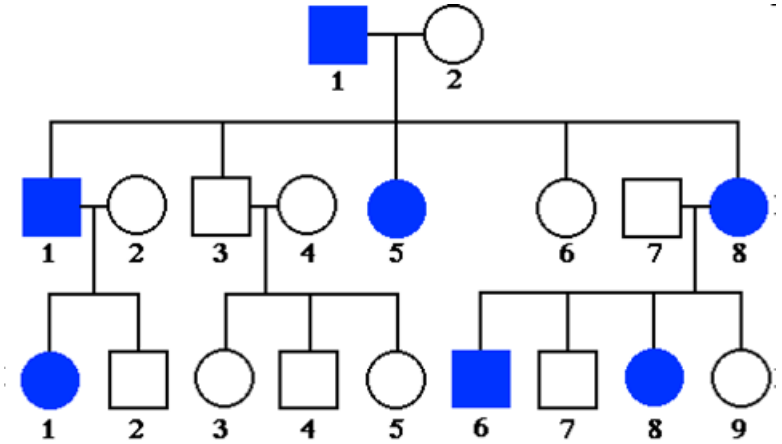
Individual	Percentage in dbSNP	Estimated total cSNPs	Estimated total non-synonymous
NA18507 (YRI)	89.1	22,727	10,261
NA18517 (YRI)	87.8	22,841	10,291
NA19129 (YRI)	87.5	22,907	10,214
NA19240 (YRI)	88.0	22,814	10,249
NA18555 (CHB)	92.8	18,722	8,447
NA18956 (JPT)	92.7	18,523	8,451
NA12156 (CEU)	94.6	18,825	8,605
NA12878 (CEU)	94.2	18,544	8,434
FSS10066 (Eur)	93.3	18,836	8,596
FSS10208 (Eur)	93.4	18,591	8,516
FSS22194 (Eur)	94.0	18,667	8,523
FSS24895 (Eur)	94.0	18,508	8,339

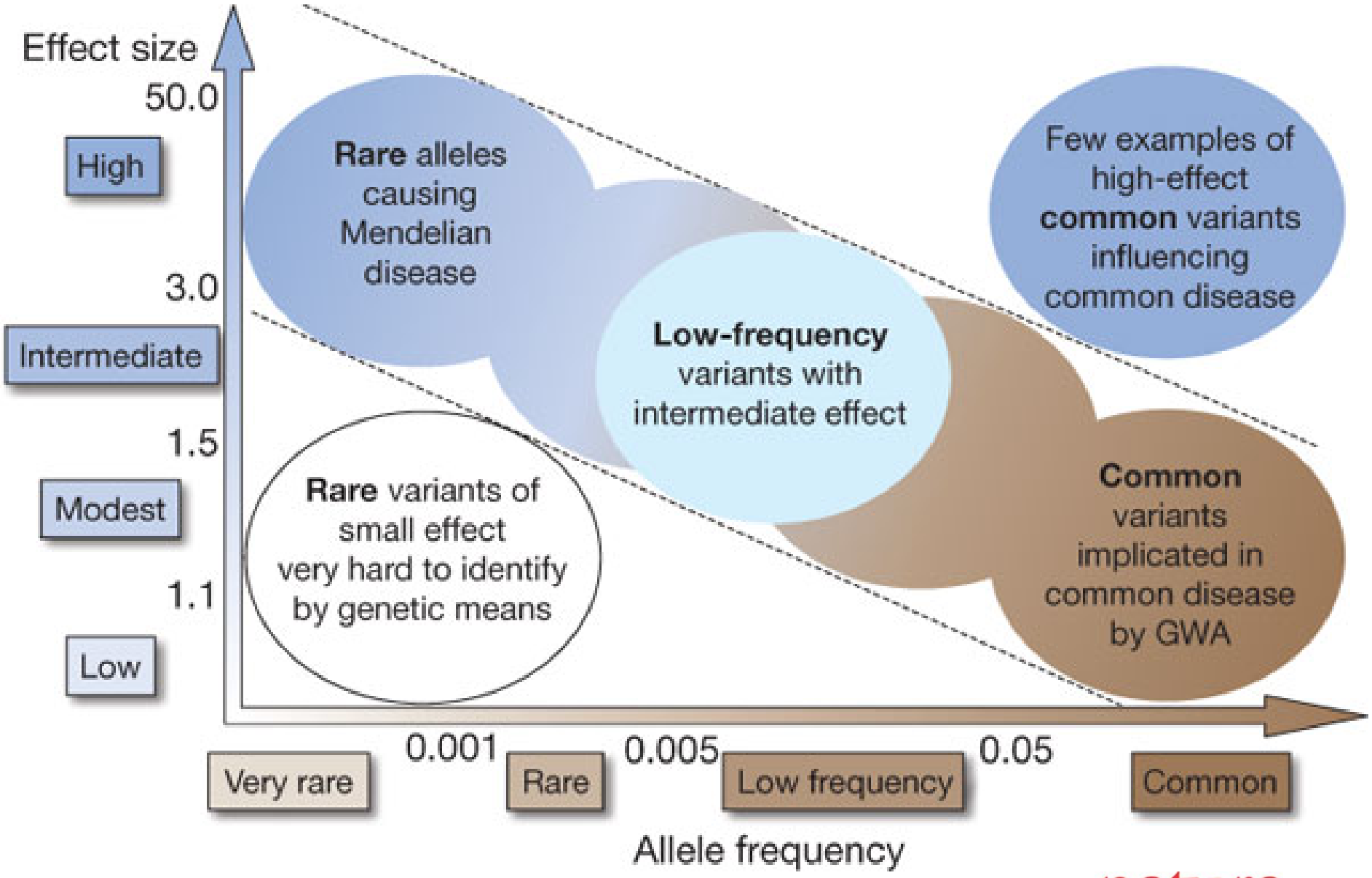
Ng et al., *Nature* (2009)

Next-generation DNA sequencing

- Sequencing of exomes & genomes has become a very practical thing to do
- **But what is it actually useful for?**







Manolio *et al.* (2009)

nature

Bad reasons to sequence exomes & genomes

- To genotype for common variants
- To profile yourself (or a patient) for GWAS risk alleles for common diseases
- much cheaper by genotyping and **not very medically informative anyways**

Good reasons to sequence exomes & genomes

- **Rare germline variants → rare diseases**
- Rare germline variants → common diseases
- Somatic mutations → cancer

- **Disease gene discovery**
- Clinical or molecular diagnostics

Mendelian disorders

- Rare, monogenic diseases
- Monogenic subsets of common diseases

- **>2,000 solved**
- **>2,000 unsolved**

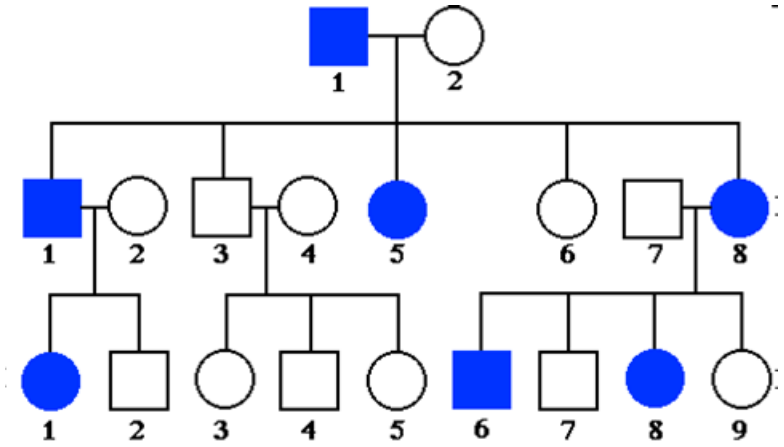


Table 2. SNPs Identified through Whole-Genome Sequencing of DNA from the Proband.*

SNP Type	No. of SNPs
Nongene	2,255,102
Gene	1,165,204
Intron	1,064,655
Promoter	60,075
3′ UTR	16,350
5′ UTR	3,517
Splice regulatory site	2,089
Splice site	112
Synonymous	9,337
Stop→stop	17
Nonsynonymous	9,069
Stop→gain	121
Stop→loss	27
Total	3,420,306

How does one pinpoint a causal variant?

Lupski et al. 2010

nature Vol 461|10 September 2009|doi:10.1038/nature08250

Targeted capture and massively parallel sequencing of 12 human exomes

Sarah B. Ng[1], Emily H. Turner[1], Peggy D. Robertson[1], Steven D. Flygare[1], Abigail W. Bigham[2], Choli Lee[1], Tristan Shaffer[1], Michelle Wong[1], Arindam Bhattacharjee[4], Evan E. Eichler[1,3], Michael Bamshad[2], Deborah A. Nickerson[1] & Jay Shendure[1]

1. Sequence exomes of several unrelated, affected individuals

2. Remove "common" variants

- 8 HapMap 'control' exomes
- public SNP databases

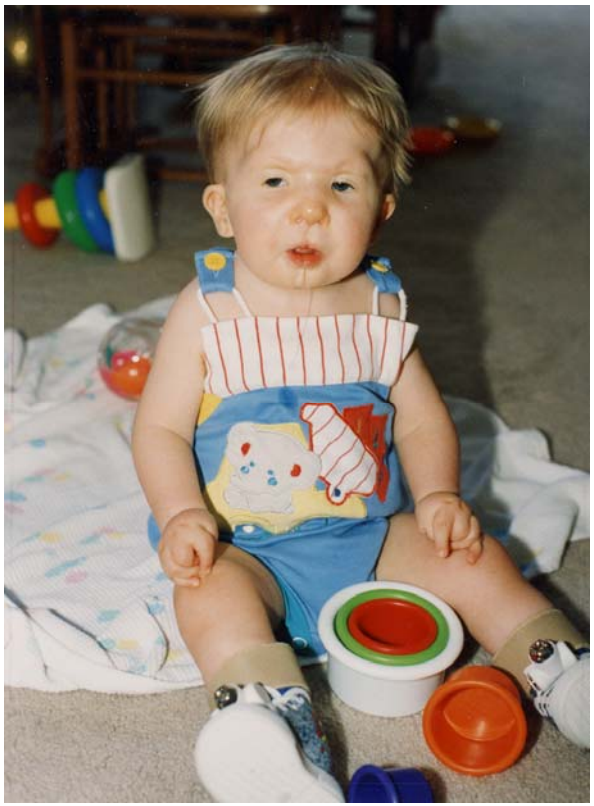
3. Find genes that contain "uncommon" variants in *all* affected individuals

Can exome sequencing be applied to identify the cause of a Mendelian disorder?

Freeman-Sheldon syndrome (FSS)

- congenital contracture syndrome
- autosomal dominant
- caused by mutations in *MYH3*

Toydemir et al., Nature Genetics (2006)



Can exome sequencing be applied to identify the cause of a Mendelian disorder?

How many genes in genome with….	**Freeman-Sheldon probands**			
	1/1	2/2	3/3	4/4
…any nsSNP, splice-site, or indel	4,510	3,284	2,765	2,479
…not in dbSNP	513	128	71	53
…not in 8 HapMap exomes	799	168	53	21
…not in either dbSNP or 8 HapMap exomes	360	38	8	**1** ***MYH3***

Ng *et al.* Nature (2009)

What about an unknown Mendelian?

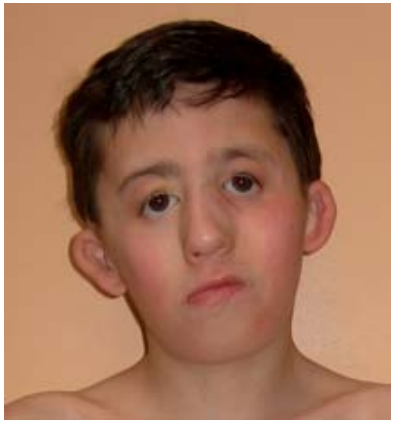
December 1979
970 The Journal of PEDIATRICS

Postaxial acrofacial dysostosis syndrome

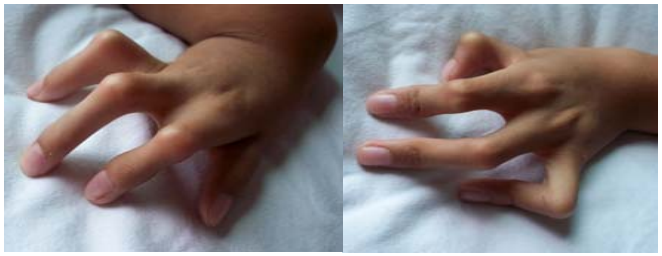
Three patients with a postaxial acrofacial dysostosis syndrome are presented; the features of these and three other previously described examples are set forth. The facies can be strikingly similar to that of the Treacher Collins syndrome. The limb deficiencies are postaxial, with absence or incomplete development of the fifth digital rays in both the upper and lower limbs. Accessory nipples have been found in most of the patients. The nature of the limb deficiencies and the accessory nipples help to distinguish this condition from Nager AFD. All of the children have normal intelligence and development; most show normal growth. All of the six cases have occurred sporadically.

Marvin Miller, M.D., *Seattle, Wash.,* **Robert Fineman, M.D., Ph.D.,**
Salt Lake City, Utah, **and David W. Smith, M.D.,*** *Seattle, Wash.*

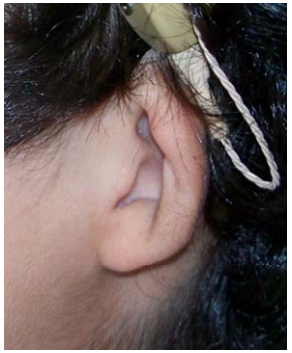
Miller syndrome



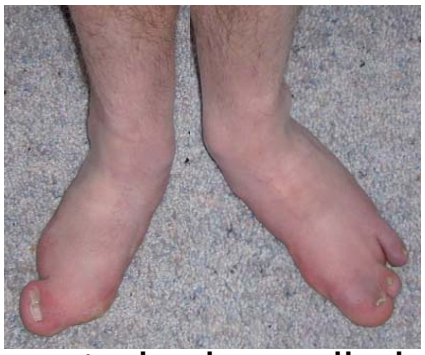
mandibular hypoplasia



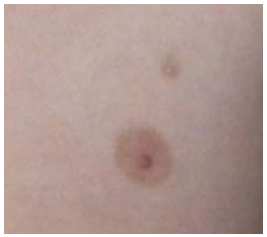
posterior upper limb defects



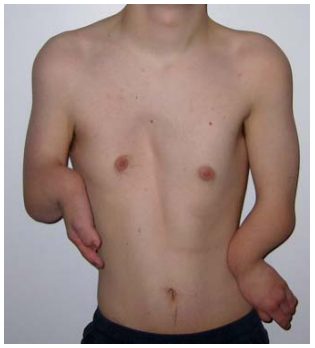
hearing loss, cupped ears



posterior lower limb defects

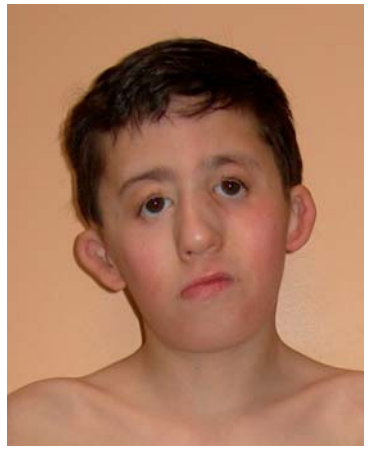


supernumerary nipples



pectus abnormalities, scoliosis

Miller syndrome



- Postaxial acrofacial dysostosis
- Presumed **autosomal recessive**
- Sequenced **4 exomes** from **3 kindreds**
- 2 siblings from kindred #1 also had CF-like lung phenotype (but not in other Miller cases)

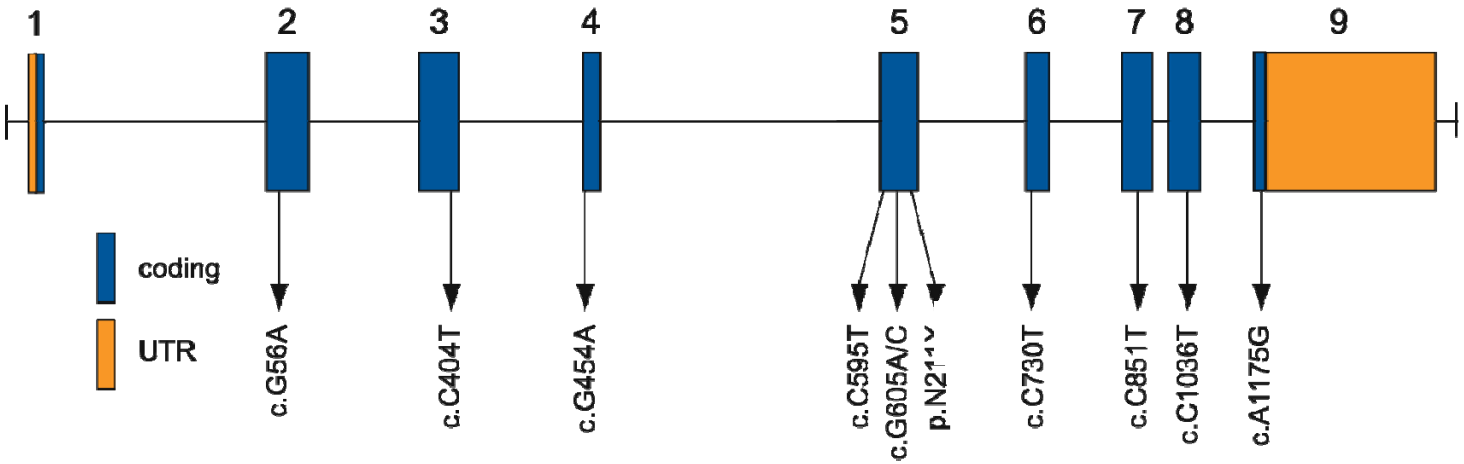
Ng *et al.* Nature Genetics (2010)

Exome analysis of Miller syndrome

How many genes in genome with…	**1-A**	**1-B**	**1-A+B**	***2/2***	***3/3***
…any **2** nsSNP, splice-site, or indel	2,789	2,777	2,278	1,740	1,461
…not in dbSNP	93	80	40	16	12
…not in 8 HapMap exomes	127	120	47	8	3
…not in either dbSNP or 8 HapMap exomes	31	26	8	**1** ***DHODH***	**1** ***DHODH***

Ng *et al.* Nature Genetics (2010)

DHODH = dihydroorotate dehydrogenase

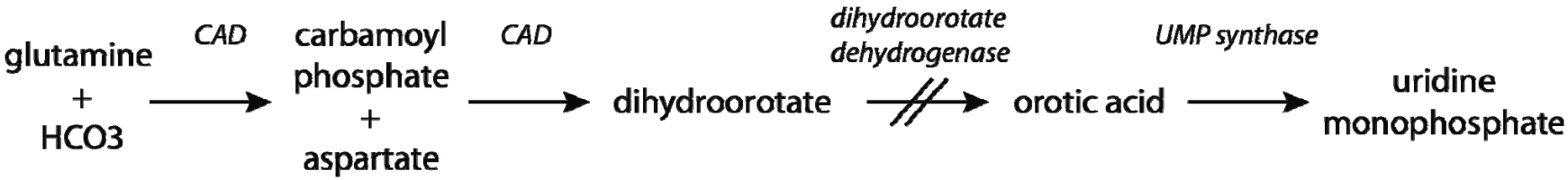


de novo pyrimidine biosynthesis

inborn error of metabolism

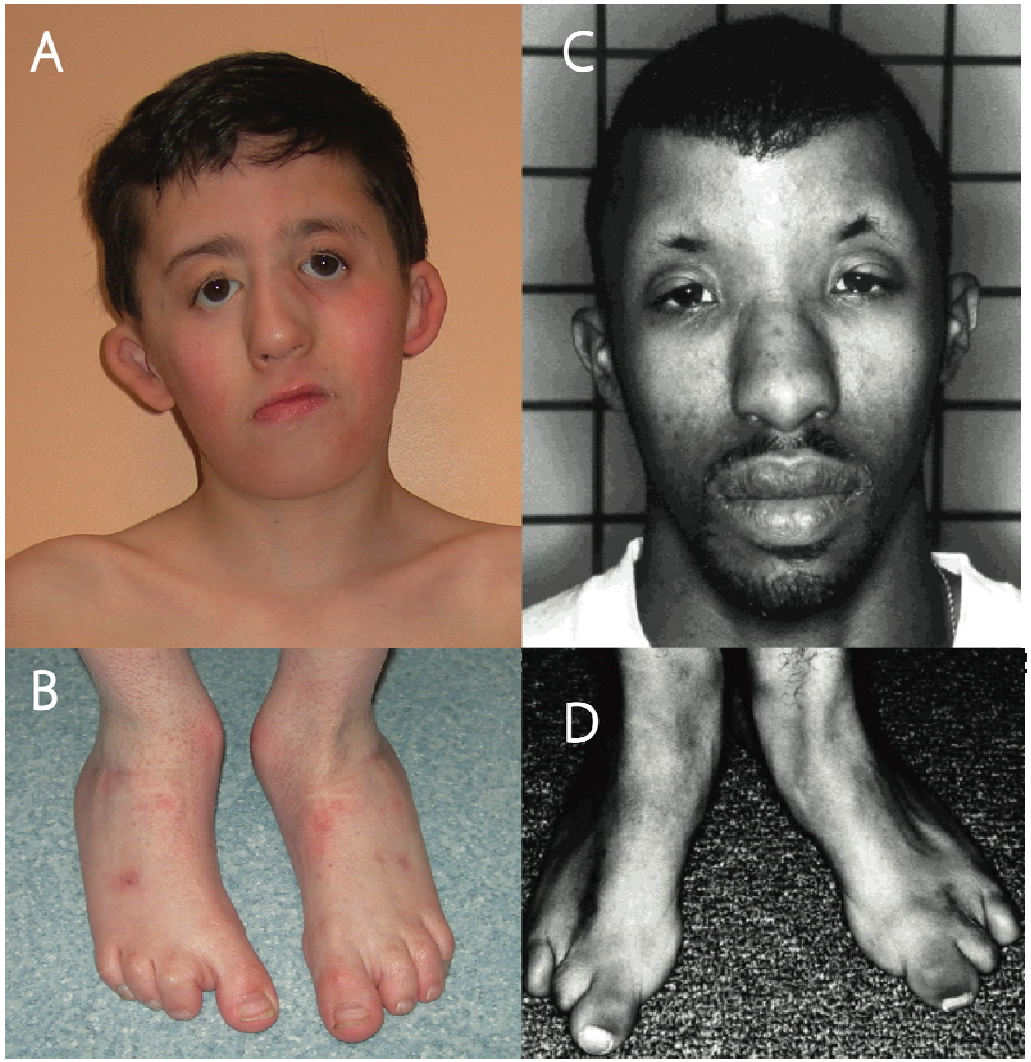
Connection to *rudimentary*

specific wing anomalies, defective oogenesis, and malformed legs
(T.H. Morgan, 1912)



Caused by mutations in *de novo* pyrimidine biosynthesis (Rawls & Fristom, 1975)

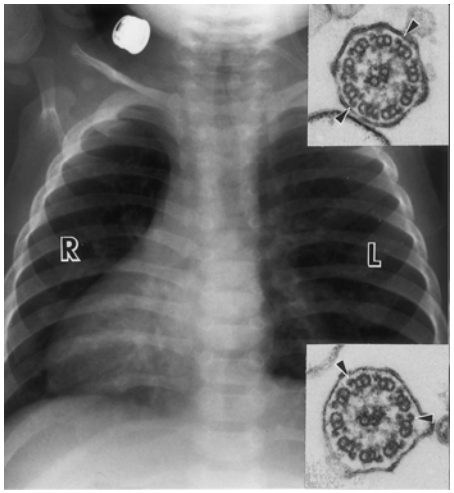
Similarity to methotrexate embryopathy



Bawle et al. *Teratology* 57:51-55 (1978)

What about the lung phenotype?

- Sibling kindred (only)
- Chronic sinopulmonary infections (CF-like)
- Compound heterozygotes for rare, damaging mutations in **DNAH5**
- Primary cilliary dyskinesia



Ng *et al.* Nature Genetics (2010)

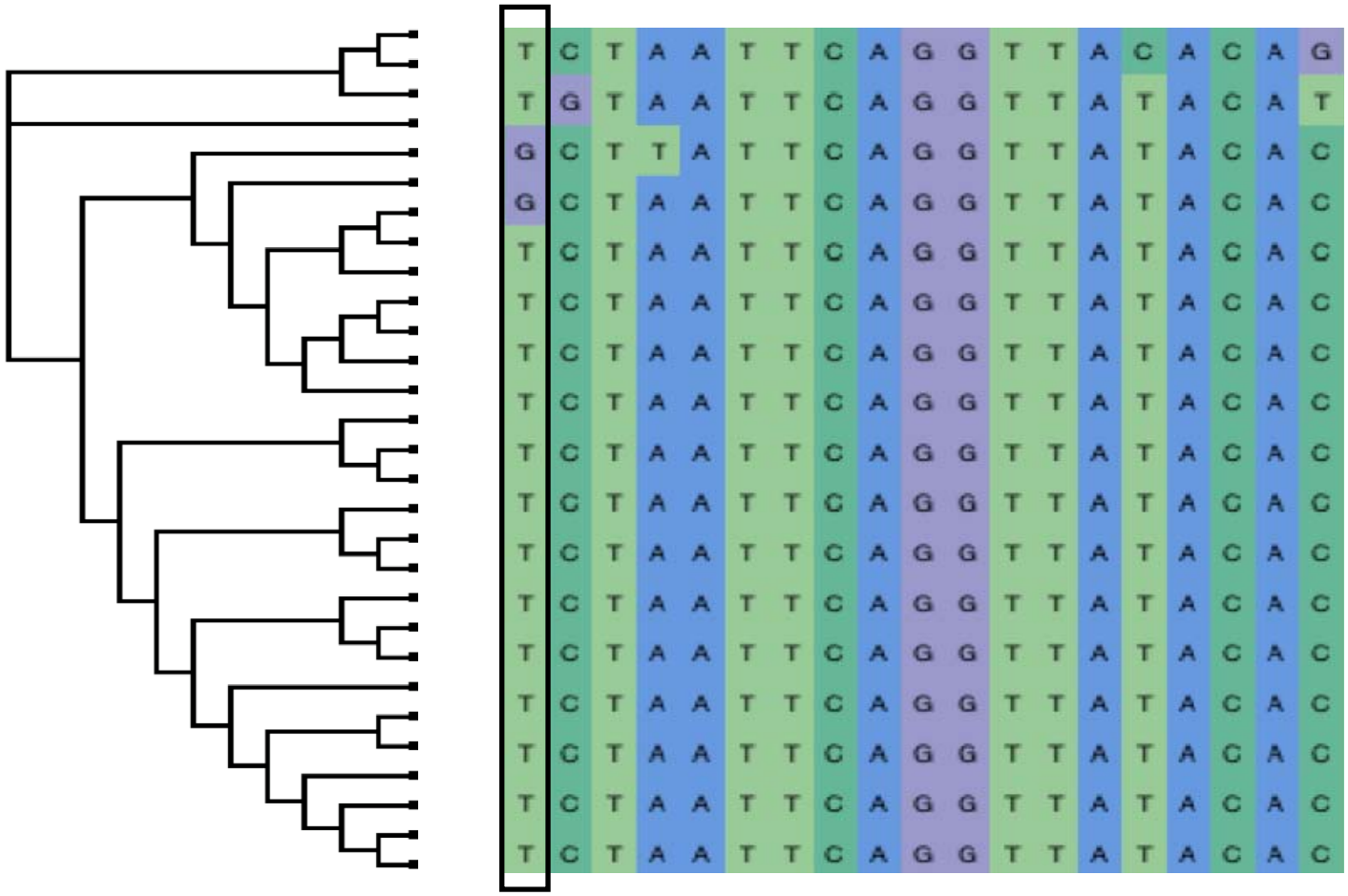
Strategies to Identify Causal Variation

- Genetics alone may be insufficient to identify causal variation
- Purifying selection (*i.e.* evolutionary 'constraint') reduces rates of evolution at functional sites

Greg Cooper

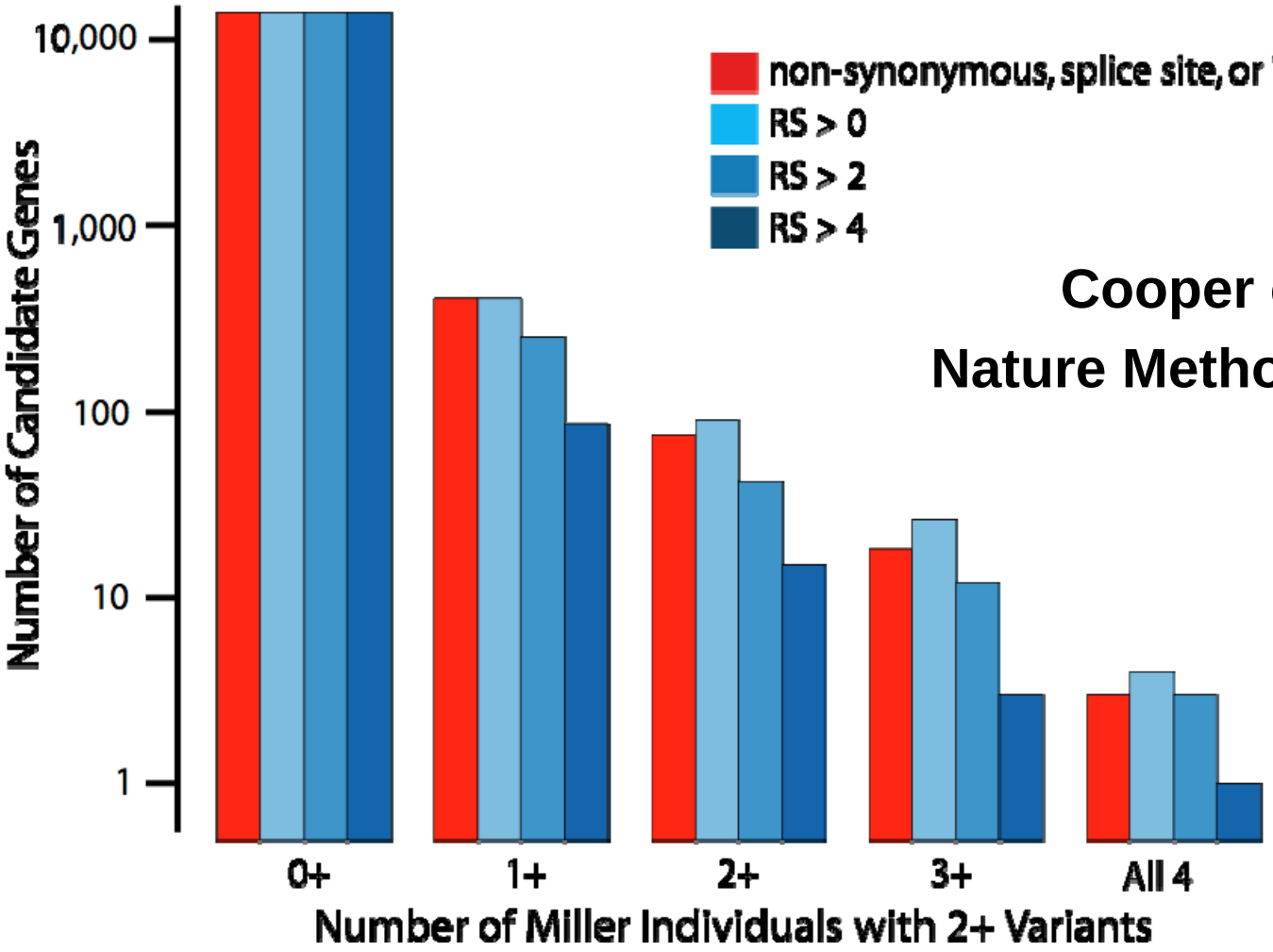
Arend Sidow

Genomic Evolutionary Rate Profiling



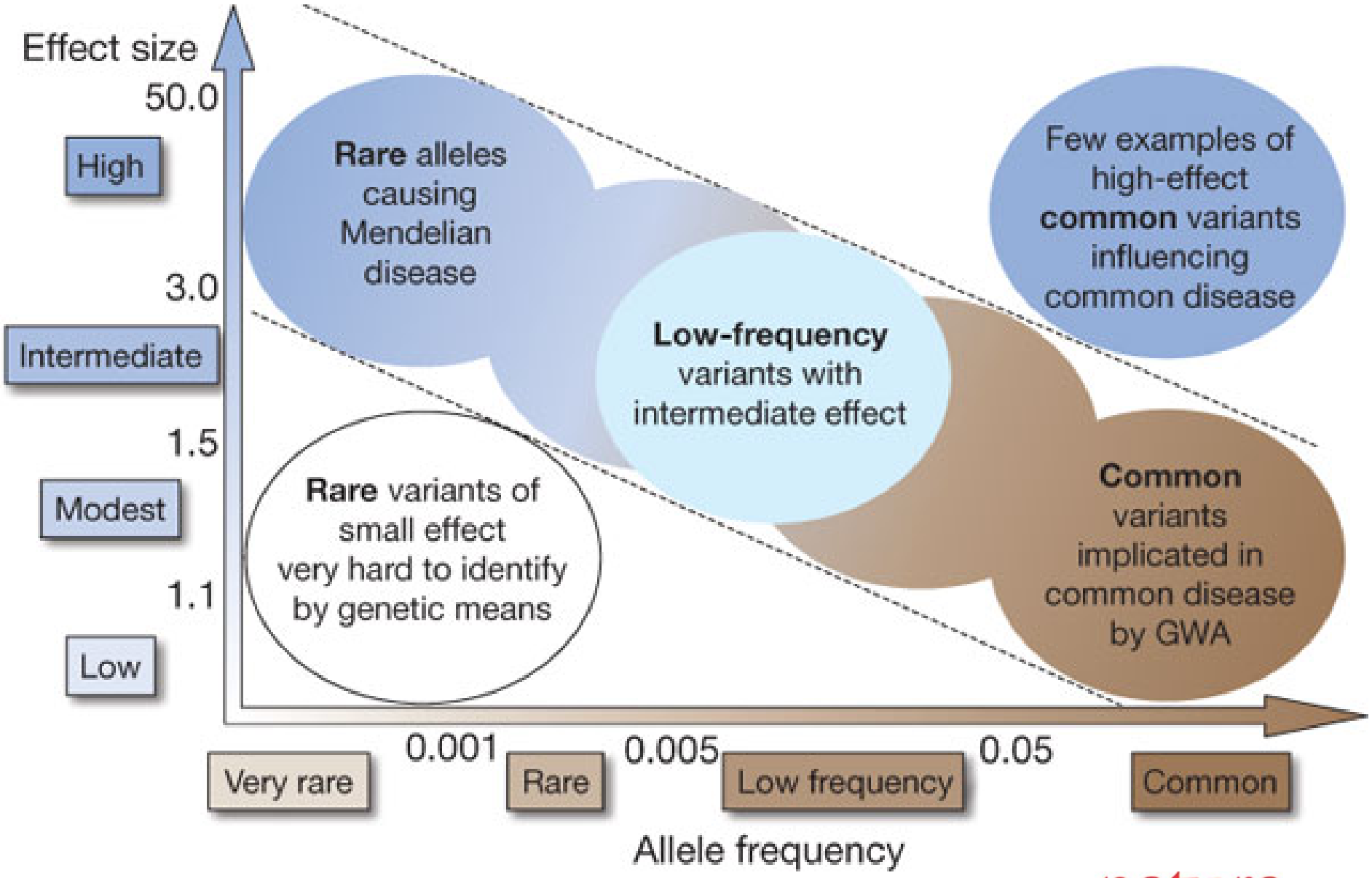
Calculate GERP score at each site based on expected versus observed # of substitutions

Constraint-Based Identification of Disease Genes



Cooper et al.
Nature Methods (2010)

DHODH rank:	226	10	2	1	1
Total (RS > 0):	13,579	397	89	26	4



Manolio *et al.* (2009)

nature

Rare, coding alleles → common diseases

- BRCA1/2 (& 11 other genes) – breast cancer
- NRXN1, SHANK3, others – autism spectrum disorders
- APP, PS1/2, UBQLN1 – early-onset Alzheimer's
- ANGPTL3/4/5, NPC1L1, PCSK9 – cholesterol, triglycerides

1) **Exome sequencing of many, many unrelated individuals, all affected by the same common disease (or phenotypic extreme)**

2) **Focused analyses on monogenic subsets of common disease**

exomes, exomes, exomes!

National Human Genome Research Institute

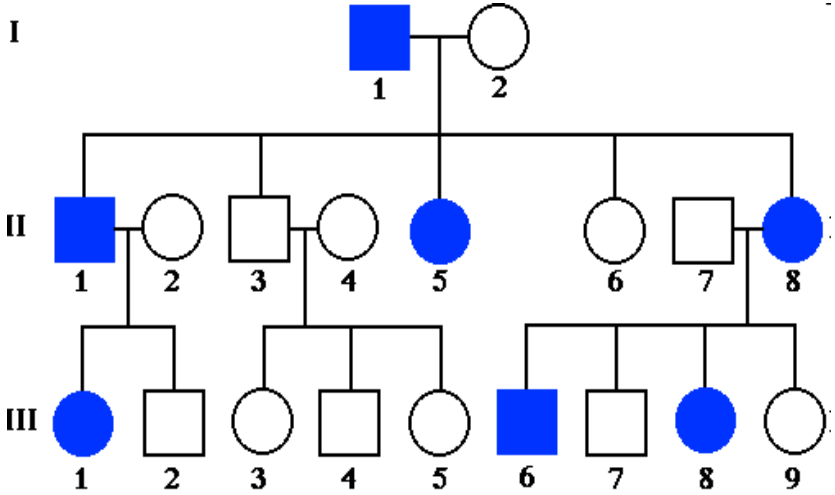
NHGRI Medical Sequencing Project

NHLBI Exome Sequecing Project ("ESP-GO")

National Heart Lung and Blood Institute
People Science Health

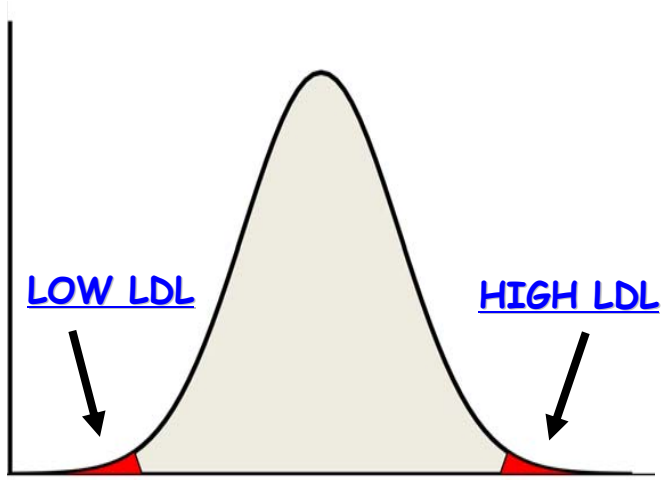
20+ mendelian diseases

~200 exomes



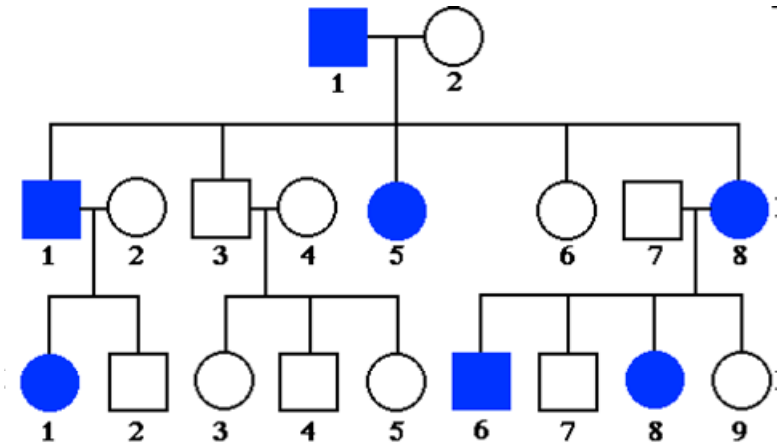
Extremes of phenotypic distribution for NHLBI-relevant traits

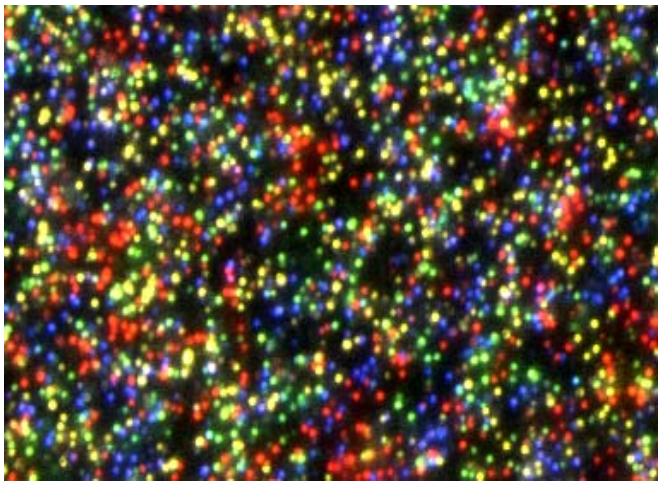
~7,000 exomes



Mendelian Genetics by Exome

- **‘High-yield’ genetics**
  - Rare, monogenic diseases
  - Monogenic subsets of common diseases
- Transition to whole genomes as costs evolve





Clinical diagnosis by exome / genome

Genetic diagnosis by whole exome capture and massively parallel DNA sequencing

Murim Choi[a], Ute I. Scholl[a], Weizhen Ji[a], Tiewen Liu[a], Irina R. Tikhonova[b], Paul Zumbo[b], Ahmet Nayir[c], Ayşin Bakkaloğlu[d], Seza Özen[d], Sami Sanjad[e], Carol Nelson-Williams[a], Anita Farhi[a], Shrikant Mane[b], and Richard P. Lifton[a,1]

[a]Department of Genetics, Howard Hughes Medical Institute, [b]Keck Foundation for Biotechnology Resources, Yale University School of Medicine, New Haven, CT 06510; [c]Department of Pediatric Nephrology, Istanbul Medical Faculty, Istanbul 34390, Turkey; [d]Department of Pediatric Nephrology and Rheumatology, Hacettepe University Faculty of Medicine, Ankara 06100, Turkey; and [e]American University of Beirut, Beirut 11072020, Lebanon

Molecular diagnosis by exome / genome

Whole-Genome Sequencing in a Patient with Charcot–Marie–Tooth Neuropathy

James R. Lupski, M.D., Ph.D., Jeffrey G. Reid, Ph.D., Claudia Gonzaga-Jauregui, B.S., David Rio Deiros, B.S., David C.Y. Chen, M.Sc., Lynne Nazareth, Ph.D., Matthew Bainbridge, M.Sc., Huyen Dinh, B.S., Chyn Jing, M.Sc., David A. Wheeler, Ph.D., Amy L. McGuire, J.D., Ph.D., Feng Zhang, Ph.D., Pawel Stankiewicz, M.D., Ph.D., John J. Halperin, M.D., Chengyong Yang, Ph.D., Curtis Gehman, Ph.D., Danwei Guo, M.Sc., Rola K. Irikat, B.S., Warren Tom, B.S., Nick J. Fantin, B.S., Donna M. Muzny, M.Sc., and Richard A. Gibbs, Ph.D.

What do we need?

- Technology is there (exomes & genomes)
- Software for sequencing is pretty much there
- **We need better tools for:**
  - **annotating variants**
  - **manipulating variant data**
  - **prioritizing variants**
    (both coding & non-coding)
  - **prioritizing candidate genes**
  - **pathway analysis for genetic heterogeneity**
  - **methods for functional analysis**

U. Washington

Anita Beck
Abby Bigham
Kati Buckingham
Zoran Brkanac
Greg Cooper
Evan Eichler
Renee George
Heidi Gildersleeve
Phil Green
Mark Hannibal
Charlie Lee
Alex McKenzie
Margaret McMillin
Brian O'Roak
Wendy Raskind
Peggy Robertson
Mark Rieder
Jerrod Schwartz
Josh Smith
Willie Swanson
Jim Thomas
Holly Tabor

Sarah Ng

Emily Turner

Debbie Nickerson

Michael Bamshad

Bamshad Lab
Nickerson / Reider Lab
Shendure Lab

Michael Knowles (UNC)
Maimoona Zariwala (UNC)
Karin Dent (U. Utah)
Ethylin Jabs (Mt. Sinai)
Koh-ichiro Yoshiura (Nagasaki Univ.)
Naomichi Matsumoto (Yokohama Univ.)
Tohru Ohta (Univ. of Hokkaido)
Norio Niikawa (Univ. of Hokkaido)
1000 Genomes Consortium

Northwest Genomics Center

National Heart Lung and Blood Institute
People Science Health

National Human Genome Research Institute

NHLBI / NHGRI Exome Project

Life Sciences Discovery Fund

Washington Research Foundation