# Some lessons from NBSeq

Steven E. Brenner
University of California, Berkeley

# Disclosures

UC Berkeley Research Agreement with Tata Consultancy Services, collaborator in this work

# Disclosures

No training in medicine, public health, ethics

No personal experience with these conditions

Not underserved by the US healthcare system

The findings and conclusions presented are those of the NBSeq authors and do not represent the official position of the California Department of Public Health Genetic Disease Screening Program, and also do not necessarily represent the official views of the National Institutes of Health. Opinions are my own

# NBSeq Collaborators


Aashish Adhikari


Yaqiong Wang

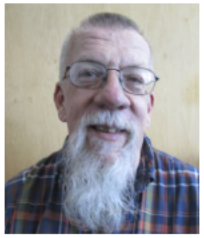

Yangyun Zou

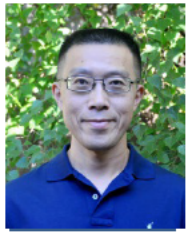

Flavia Chen


Jennifer Puck, PI


Pui Kwok, PI


Barbara Koenig, PI


Bob Currier


Hao Tang


Neil Risch


Joseph Sheih


Bob Nussbaum


Rajgopal Srinivasan


Uma Sunderam


Renata Gallagher


Sean Mooney

Beatriz Anguiano, Laia Bars, Amy Brower, Ajithavalli Chellappan, Binghuang Cai, Brad Dispensa, Carol Fraser-Browne, George Freedman, Julie Harris-Wai, Josie Johnston, Galen Joseph, Mark Kvale, Jamie King, Richard Lao, Dae Lee, Jeremy Michelson, Alan Nguyen, Erik Parens, Monica Smith, Dedeepya Vaka, Eunice Wan, Rachel Zacharias, Brandon Zerbe
Joy Boyer, Melissa Parisi, Tiina Urv, Anastasia Wise

UCSF University of California San Francisco

Berkeley UNIVERSITY OF CALIFORNIA

UW Medicine UW SCHOOL OF MEDICINE BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

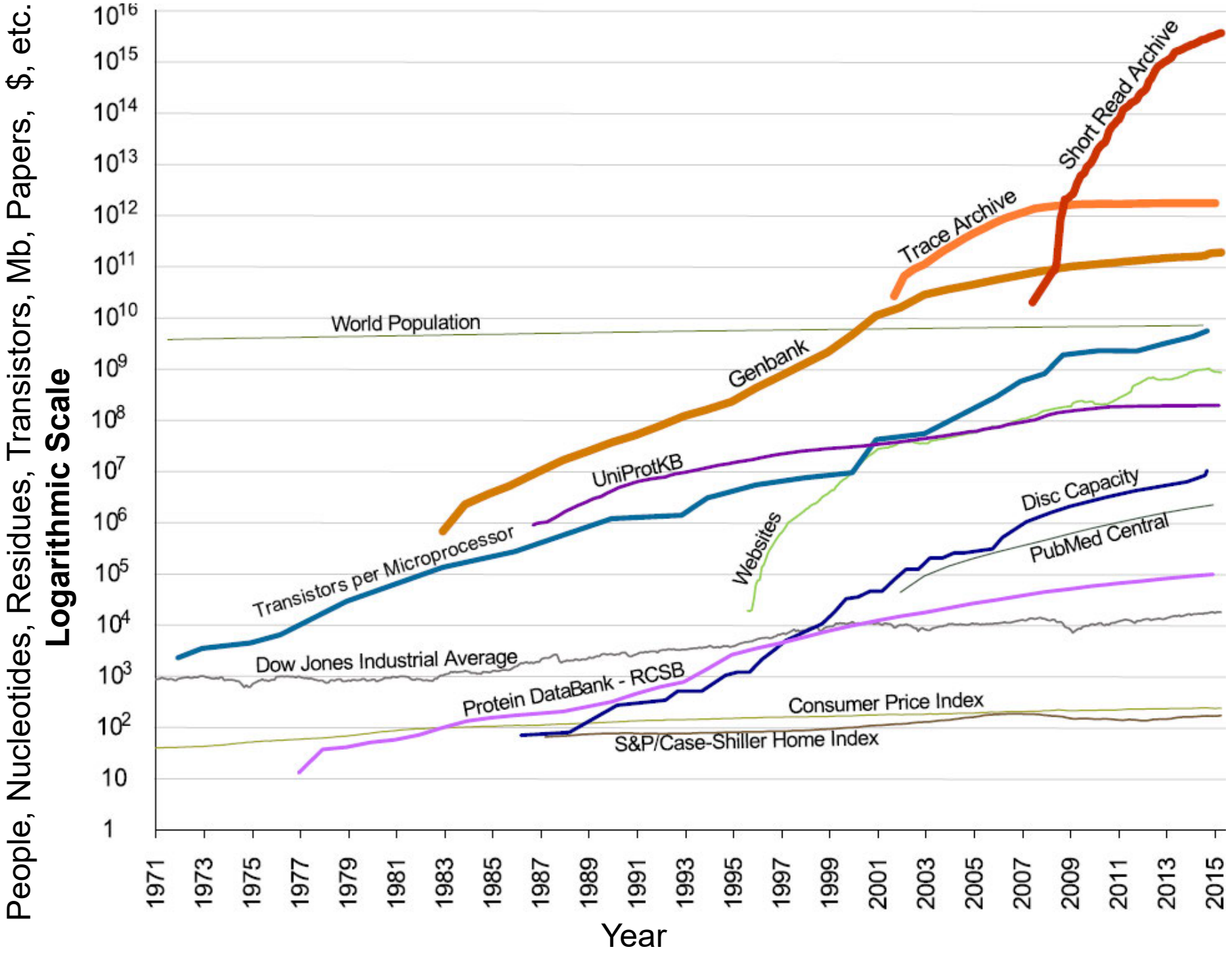
TATA CONSULTANCY SERVICES

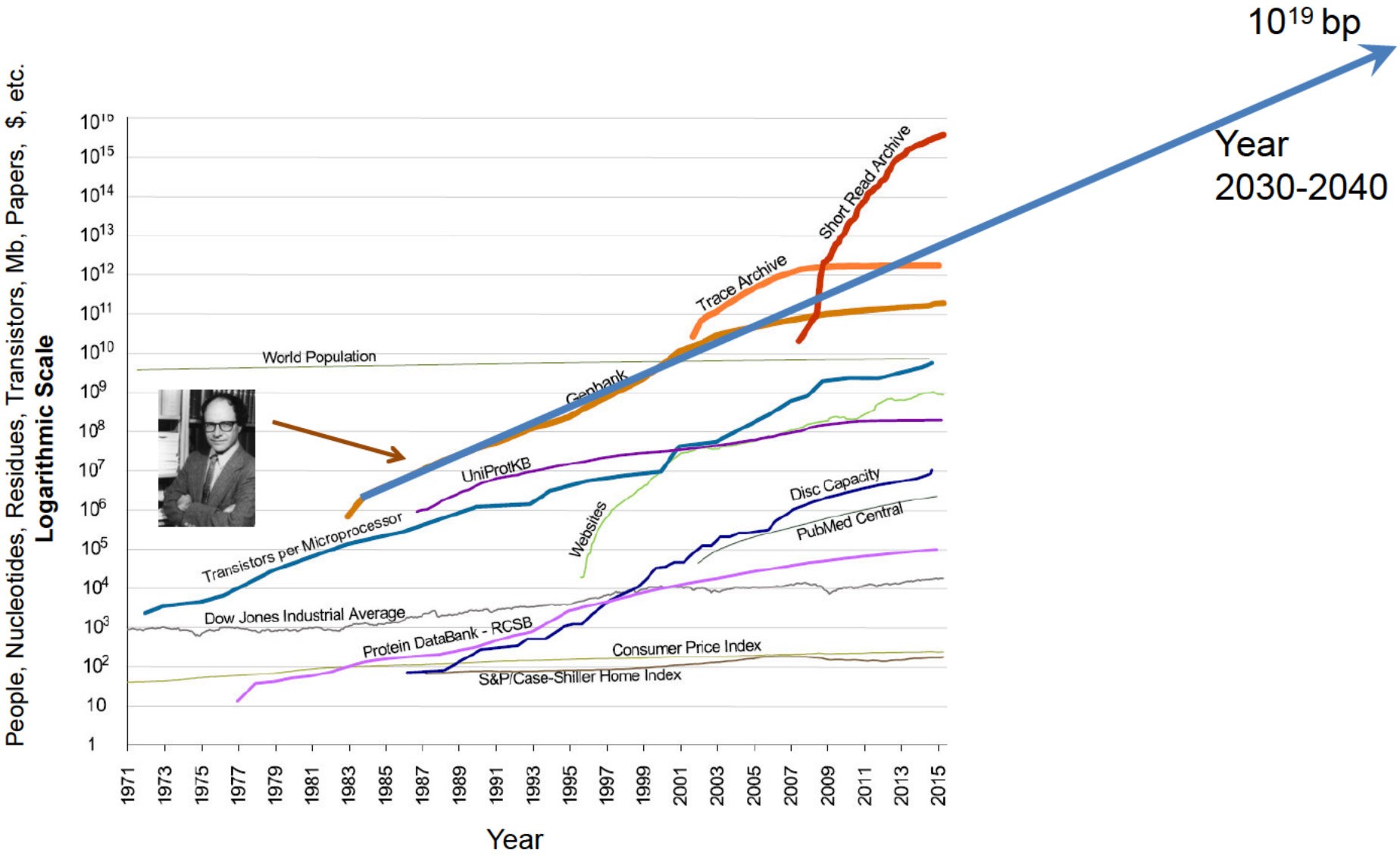
CONSORTIUM ON LAW, SCIENCE & HEALTH POLICY

NIH *Eunice Kennedy Shriver* National Institute of Child Health and Human Development

National Human Genome Research Institute

The findings and conclusions presented are those of the authors and do not represent the official position of the California Department of Public Health Genetic Disease Screening Program, and also do not necessarily represent the official views of the National Institutes of Health


People, Nucleotides, Residues, Transistors, Mb, Papers, $, etc.
Logarithmic Scale
$10^{16}$
$10^{15}$
$10^{14}$
$10^{13}$
$10^{12}$
$10^{11}$
$10^{10}$
$10^{9}$
$10^{8}$
$10^{7}$
$10^{6}$
$10^{5}$
$10^{4}$
$10^{3}$
$10^{2}$
10
1
1971
1973
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
2015
Year
Short Read Archive
Trace Archive
World Population
Genbank
UniProtKB
Transistors per Microprocessor
Websites
Disc Capacity
PubMed Central
Dow Jones Industrial Average
Protein DataBank - RCSB
Consumer Price Index
S&P/Case-Shiller Home Index


People, Nucleotides, Residues, Transistors, Mb, Papers, $, etc.
Logarithmic Scale
Year
World Population
Genbank
Trace Archive
Short Read Archive
UniProtKB
Transistors per Microprocessor
Websites
Disc Capacity
PubMed Central
Dow Jones Industrial Average
Protein DataBank - RCSB
Consumer Price Index
S&P/Case-Shiller Home Index
$10^{19}$ bp
Year
2030-2040

COMMENTARY

# Common sense for our genomes

A personal DNA sequence is not yet practically useful. But it could be, argues **Steven E. Brenner**, if we had the right resources available to interpret genomes.

Revelation of the complete DNA sequences of James Watson and J. Craig Venter elicited headlines in recent months, but most press reports struggled to offer meaningful interpretations. The most noted observation was that Venter has a particular gene variant predisposing him to cardiac disease, although his family history was enough to let him know about this general risk. If the genome is so revealing, why was so little revealed?

It is telling that Venter said he learned about the cardiac disease gene in a newspaper report. Put simply, even we in the scientific community can't easily come to grips with what we know. The effects of gene variations are scattered in hundreds of databases, across hundreds of interpretative reports in clinical laboratories, and among millions of manuscripts and patent applications. And although some papers

unique single-nucleotide substitution[4].

Visionary geneticists have long contemplated building a resource to consolidate our understanding of genome variation. However, academic squabbles and misunderstandings caused the most comprehensive effort — involving hundreds of scientists backed with millions of dollars — to founder[5]. Perhaps they were premature? Until recently, it was rarely productive to look beyond a single gene known to be of research or clinical interest. Today, the situation has changed radically. With the prospect of inexpensive personal genome sequences, there is profound impetus for integrating our knowledge of genetic variation and its effect on a genomic scale.

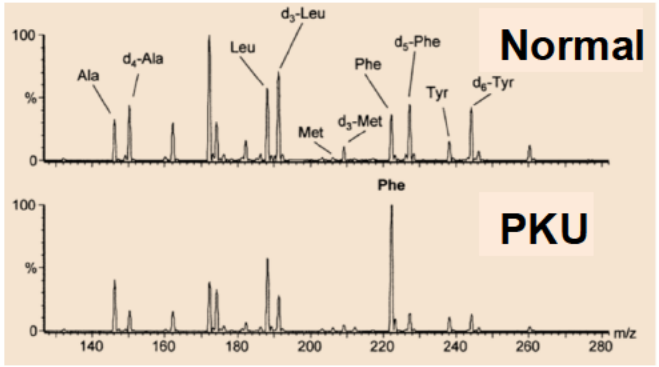
## Covering the bases

Many of the foundations for describing human

# Current newborn screening for metabolic disorders is based on blood biochemical metabolites

**Dried blood spots**


**Tandem Mass Spectrometry (MS/MS) biochemical profile**




**Positive screens lead to follow-up and diagnosis in metabolic centers**

**~500,000**
babies per year born and screened in California

**48**
types of metabolic disorders are screened by MS/MS

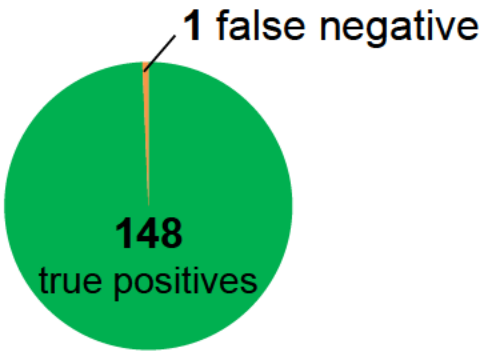
sensitivity > 99%
specificity > 99%

**~150**
newborns per year are diagnosed with some metabolic disorder

# Current status of MS/MS newborn screening

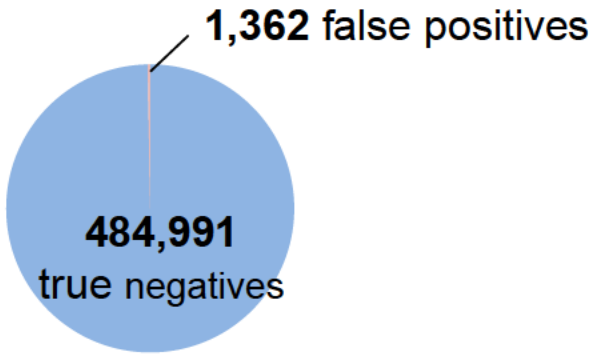
**MS/MS Screening (CDPH 2015)**

**Sensitivity: 99.3%**




149 affected newborns

**Specificity: 99.7%**




486,353 unaffected newborns

## Implications of False Negatives

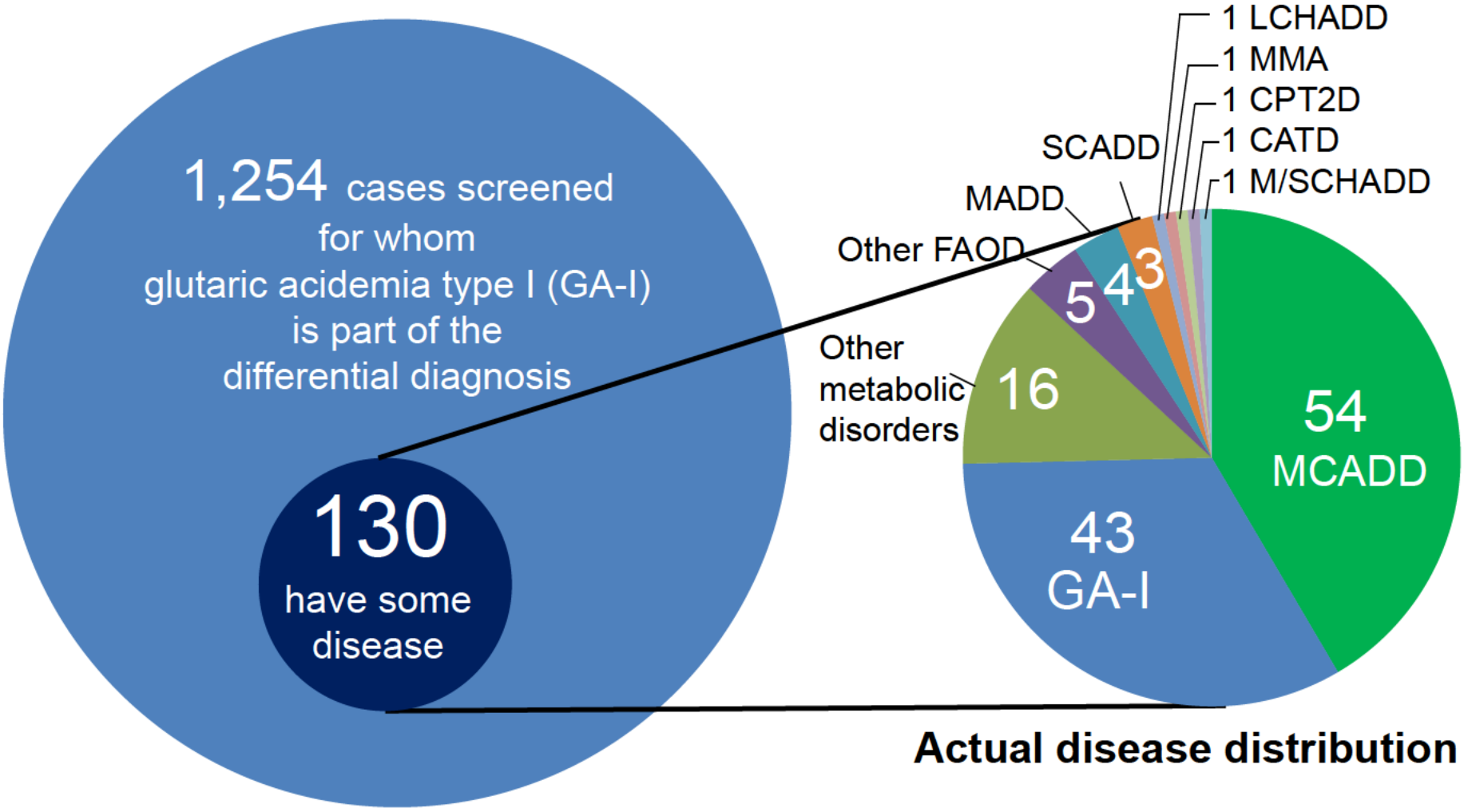
Severe neurological symptoms,
irreversible damage,
death

## Implications of False Positives

Medical care to resolve false positive
Parental stress and anxiety,
"fragile baby syndrome,"
reduced parent bonding

# MS/MS has specificity of 99.5%, yet for some conditions PPV is low

## The same MS/MS screen can sometimes resolve to different disorders




CPDH data from 7 July 2005 to 31 December 2013

# NBSeq U19 HD077627 Grant Abstract

Newborn screening (NBS) is an essential public health program in all 50 states. The falling cost of whole genome/ exome sequencing provides an opportunity to ask whether whole genome analysis (WGA) might serve as a method of cost-effective newborn screening for any and every condition. We will address certain critical questions raised by the application of this technology to NBS. We will use Whole Exome Sequencing (WES) as a cost-effective method of WGA in 1620 newborn blood spots that are linked to the clinical data of the newborns. We will then test WES as a NBS Tool for metabolic and immunological disorders. These data will be used to:

## 1) compare the sensitivity and specificity of mutation data with biochemical testing,

## 2) identify gene variants that predict which children with certain metabolic disorders are at greater risk for metabolic decompensation

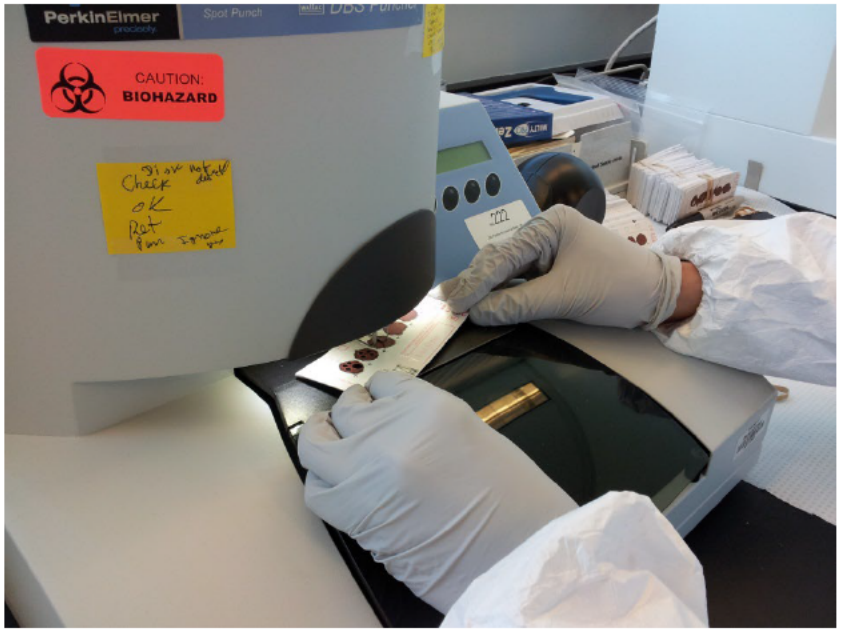
3) identify mutations in genes responsible for those primary immunodeficiencies that are not detected by the current T-Cell receptor excision circle assay used for severe combined immunodeficiency screening, and

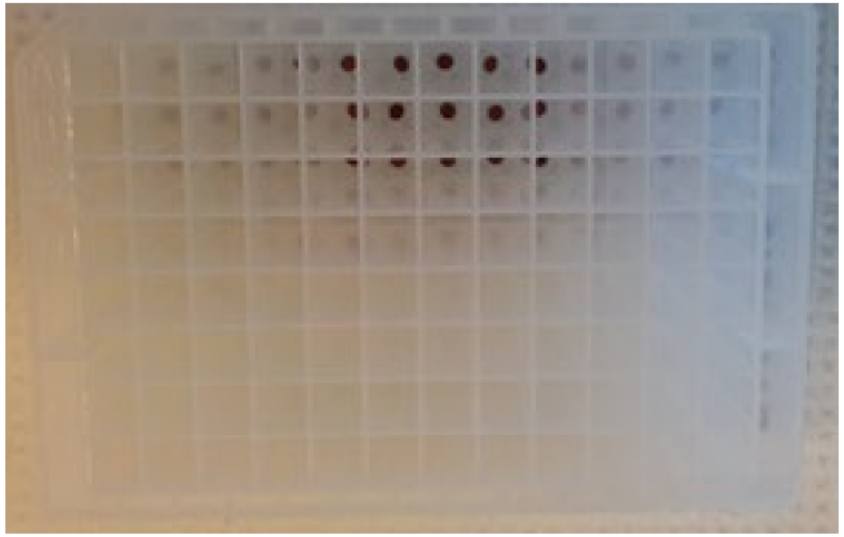
4) scan 9 genes for variants that are clinically important for drug metabolism and would be typical "secondary findings" if WES were to be used as a NBS method.

We will also develop a participant protection framework for conducting WGA during the neonatal period, determine the views, perspectives, and value preferences of key stakeholders about using WGA for NBS, collaborate with the UC Hastings Consortium on Law, Science and Health Policy, to identify the legal and constitutional issues for using WGA, and for incorporating PGx into NBS programs, and develop and disseminate policy recommendations for expanded NBS programs based on WGA.

**Public Health Relevance**

We plan to test the use of one method of whole genome analysis, whole exome sequencing, as the sole method of newborn screening for disorders currently being screened for and others that are missed by current newborn screening but for which newborn screening might be appropriate. This would provide substantial public health benefit for newborns who might otherwise go undiagnosed and untreated while serious, even irreversible, damage occurs. Equally important, we will explore parents' interest in incidental findings and develop appropriate legal frameworks to make sure NBS by WGA can be done ethically.



Carol Fraser-Browne

California Department of PublicHealth

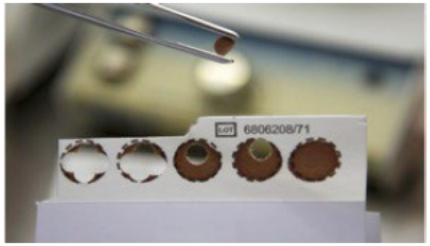
UCSF
University of California
San Francisco

# NBSeq

# Whole exome sequencing of dried blood spots from most affected children for 8.5 years in CA

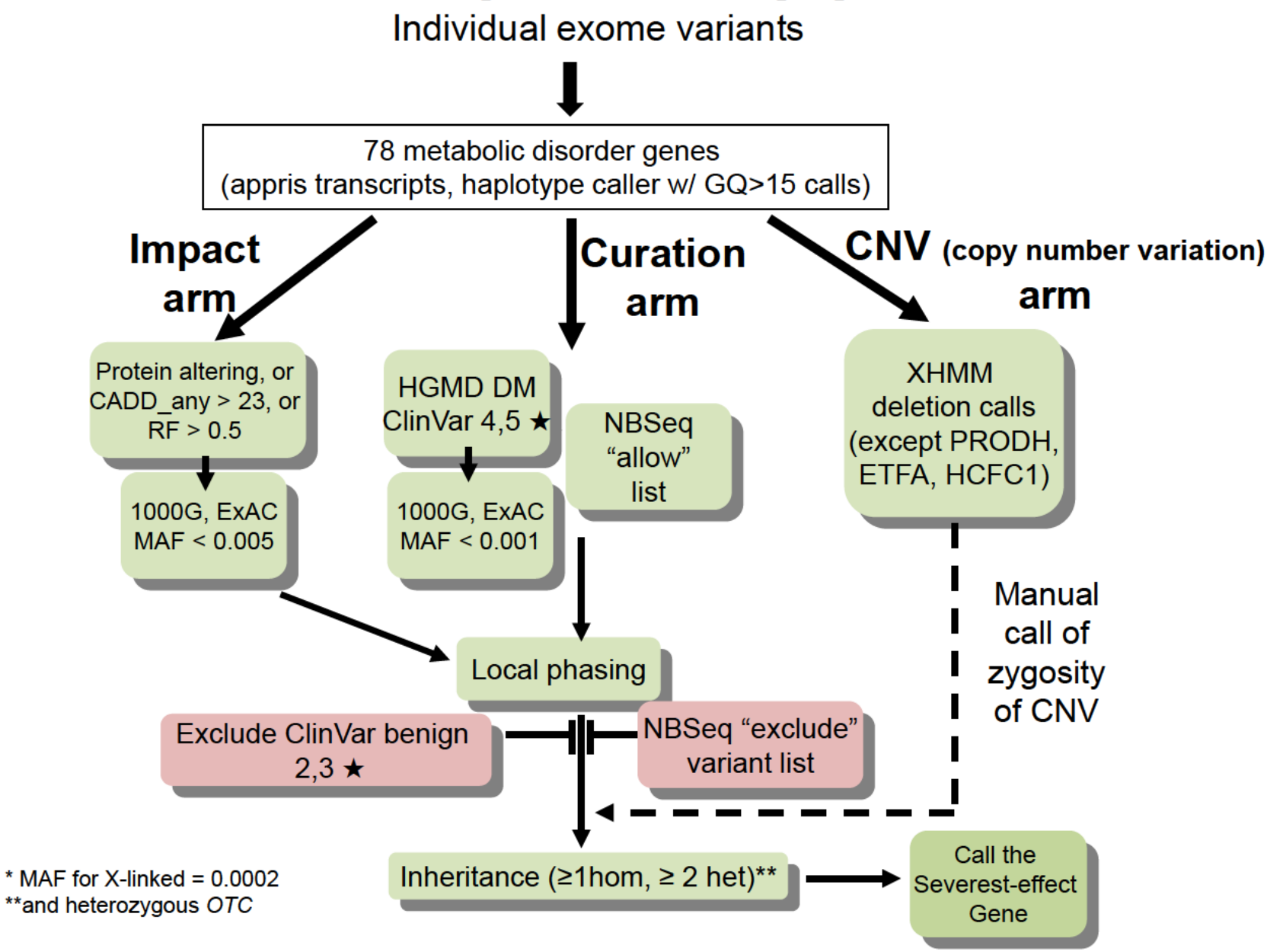
| Actual metabolic disease state | MS/MS screening: Positive | MS/MS screening: Negative |
|---|---|---|
| Affected | **MS/MS True Positives**<br>1334 DBS | **MS/MS False Negatives**<br>13 DBS |
| | **900 exomes** | |
| Healthy | **MS/MS False Positives**<br>10,011 DBS | **MS/MS True Negatives**<br>4.44 million DBS |
| | **316 exomes** | **0** |

CPDH data from 7 July 2005 to 31 December 2013



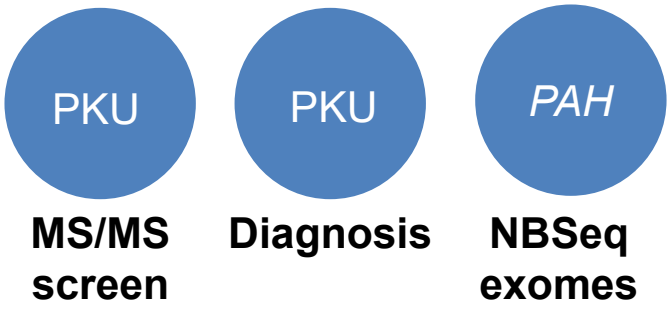
The de-identified residual DBS from the California Biobank for this project (SIS request number 496) were obtained with a waiver of consent from the Committee for the Protection of Human Subjects of the State of California, under project no. 14-07-1650 and in compliance with CDPH Biospecimen /Data Use and Confidentiality Agreement.

# Variant interpretation pipeline



# Example of a typical true positive case: Two variants flagged in relevant gene

PKU is caused by defects in the *PAH* gene




**All variants in the PAH gene not flagged by the pipeline in the exome**

| Mutation Impact | 1000 Genomes MAF | CADD | HGMD | ClinVar | Genotype |
|---|---|---|---|---|---|
| Nonsynonymous (L348V) | 0.00 | 24.7 | DM | Pathogenic(★) | Heterozygous |
| Nonsynonymous (R241H) | 0.00 | 31 | DM | Pathogenic(★) | Heterozygous |

→ **Two variants flagged by the pipeline**

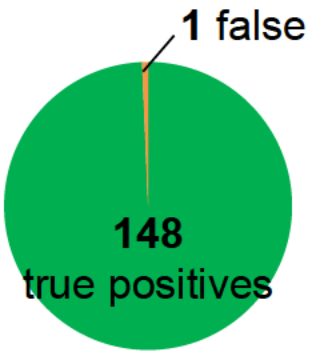
**Both are compelling:**

- alter protein sequence
- very rare
- curated as pathogenic
- predicted as damaging

# Exomes unsuitable alone for newborn screening for IEMs

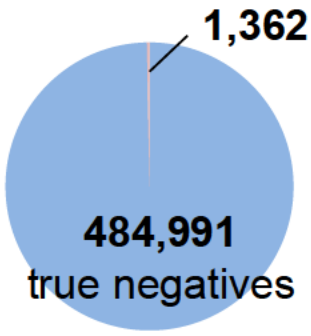
**MS/MS Screening (CDPH 2015)**

**Sensitivity: 99.3%**




149 affected newborns

**Specificity: 99.7%**



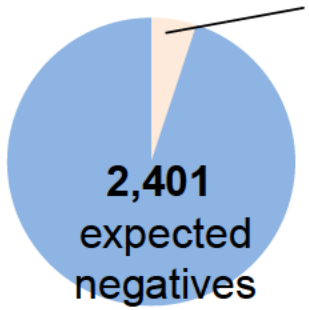

486,353 unaffected newborns

**NBSeq exomes**

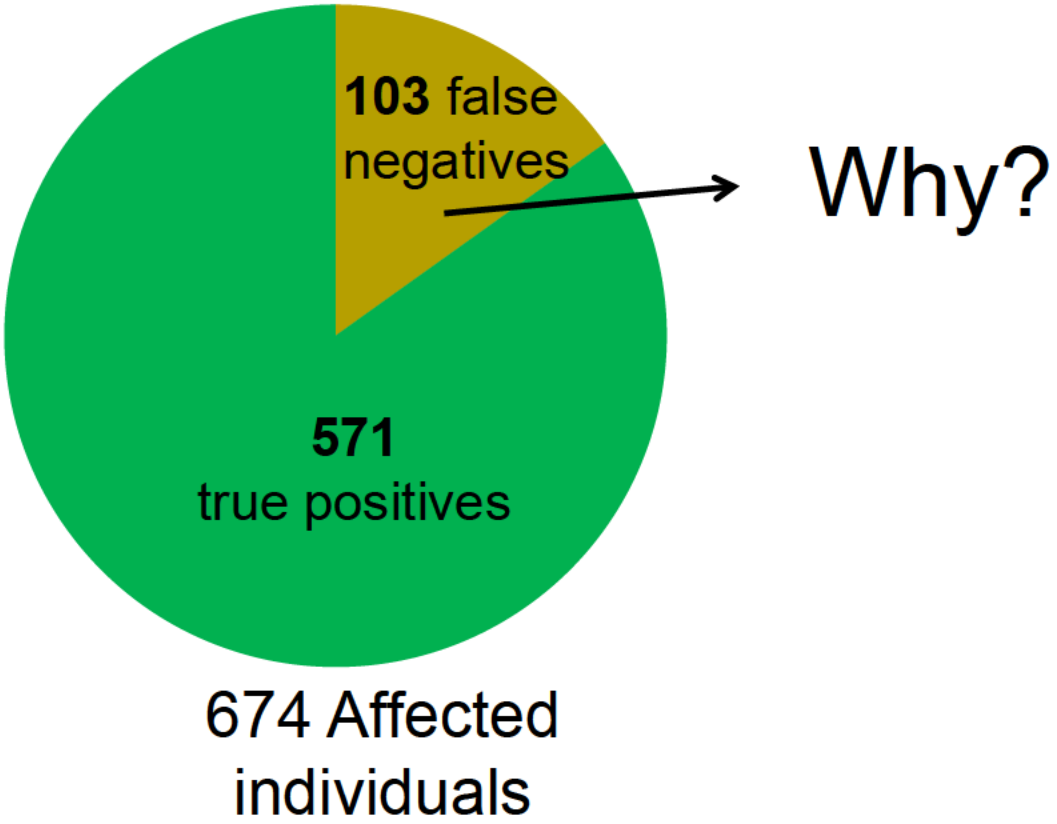
**Sensitivity: 88%**


674 Affected individuals

**Specificity: 95%**

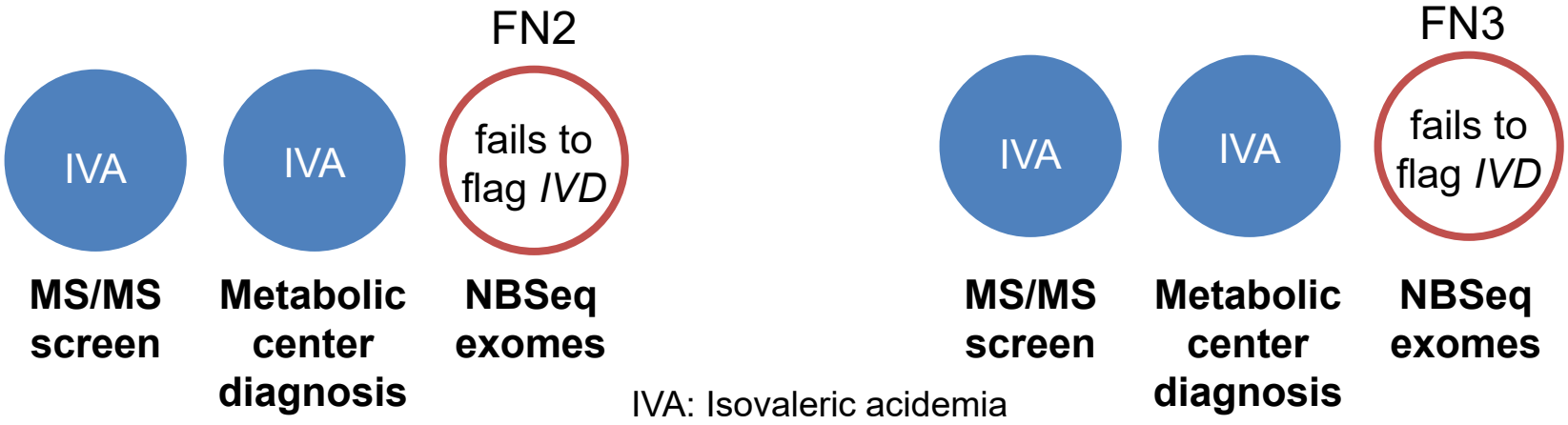



2504 individuals (1000 Genomes Phase 3)




**Half lack any missense variants in relevant genes**
**Half have only one heterozygous missense variant**

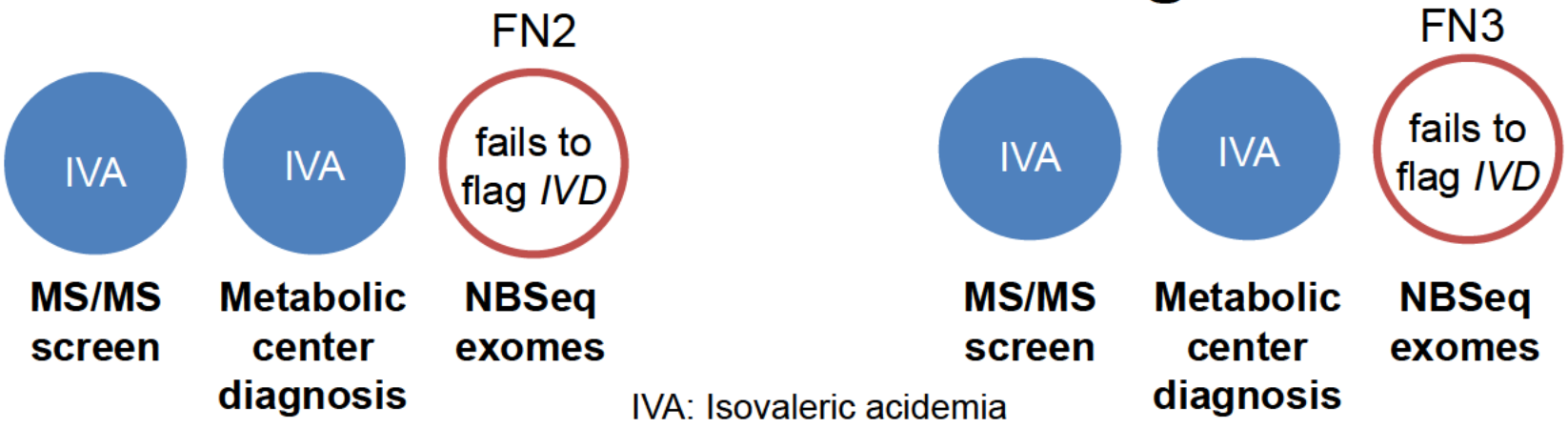
# Some sources for false negatives



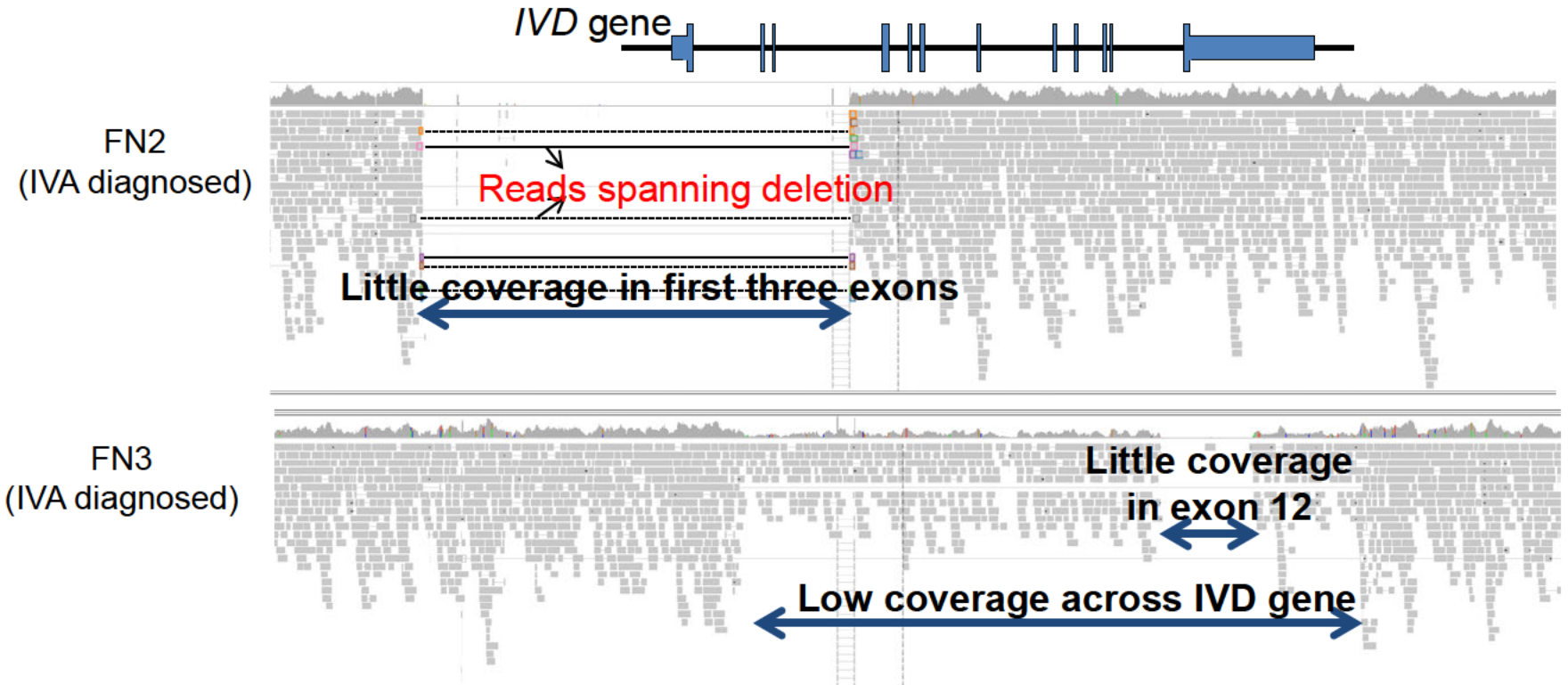

IVA: Isovaleric acidemia

**No interesting variants in the IVD gene in the exome**

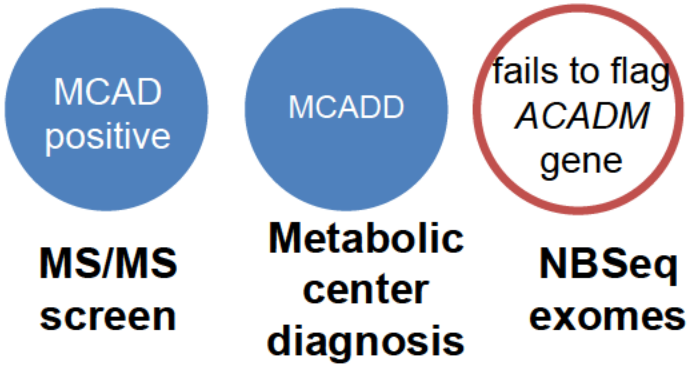
# Some sources for false negatives




## Whole genome sequencing reveals potential *IVD* deletions

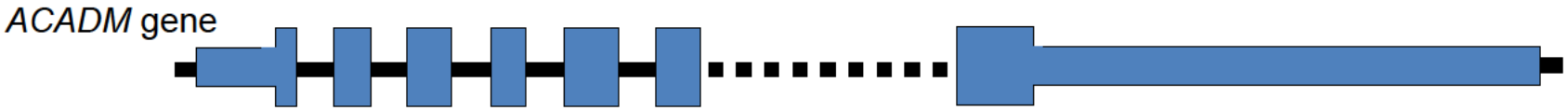


# Some sources for false negatives



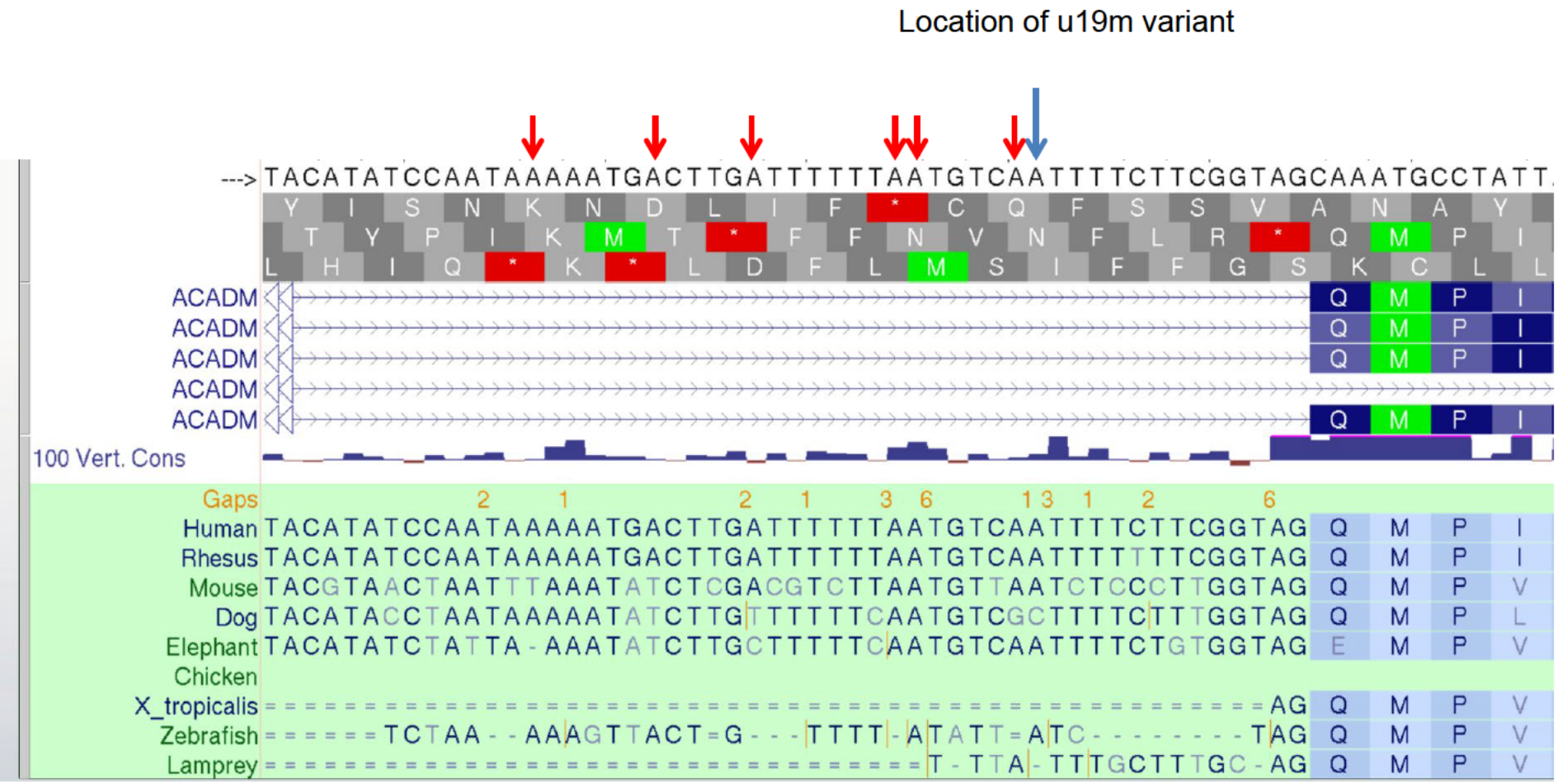

## Noncoding variants could be pathogenic

Example: MCADD (Medium-chain acyl-CoA dehydrogenase deficiency)




| | Variant 1(flagged) | Variant 2 (not flagged) |
|---|---|---|
| **Mutation impact** | Nonsynonymous (Y67H) | Intronic(-14A>G) |
| **MAF(1000 Genomes)** | 0.0000 | 0.0002 |
| **HGMD** | DM? | - |
| **ClinVar** | 5 (★★) | - |
| **Genotype** | Heterozygous | Heterozygous |

# Branch point prediction does not implicate the variant




↓ Predicted candidate branch points (SVM-BPfinder[1])

1.http://regulatorygenomics.upf.edu/Software/SVM_BP/ - accessed 11 Nov 2016

# Exome identifies a potential late-onset disorder

VLCAD MCAD

**MS/MS screen**

VLCADD Carrier

**Metabolic center diagnosis**

GA-II (*ETFDH* gene)

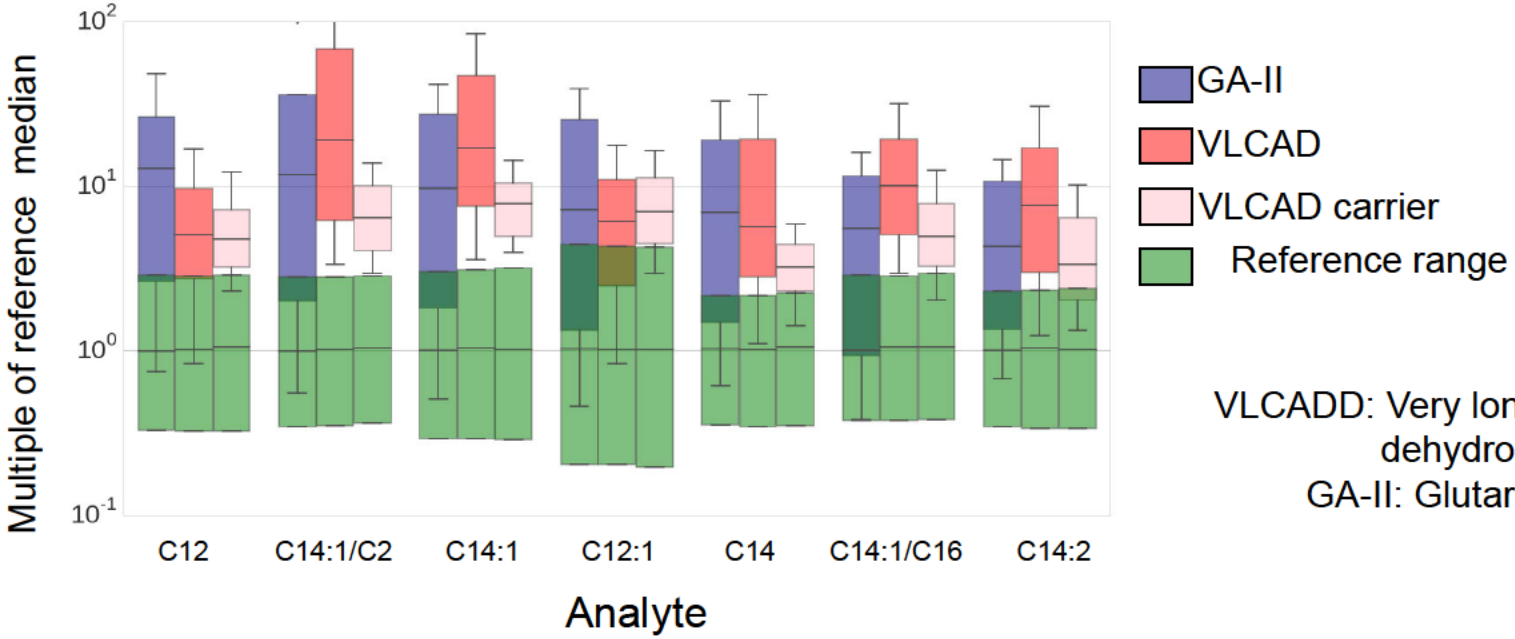
**NBSeq exomes**

Exome flagged two heterozygous variants in *ETFDH*

**Mutations in *ETFDH* cause GA-II**

| Mutation | A84T | R175H |
|---|---|---|
| Mutation impact | Nonsynonymous | Nonsynonymous |
| MAF(1000 Genomes) | 0.0002 | 0 |
| HGMD | DM | DM |
| ClinVar | 5 (★) | 5 (★) |
| Genotype | Heterozygous | Heterozygous |

Identical pair of compound heterozygous mutations reported in 6 *late-onset* GA-II patients from 5 families[1]




VLCADD: Very long chain acyl-Co-A dehydrogenase deficiency
GA-II: Glutaric acidemia type II

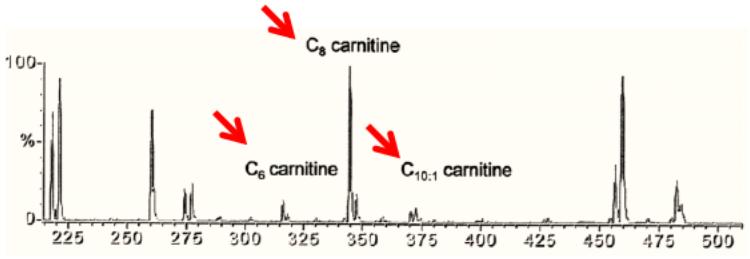


1. Wang (2011) *J Mol Med* **89:** 569 DOI: 10.1007/s00109-011-0725-7

# Current potential for sequencing in screening

## Preliminary findings

https://sites.psu.edu/siowfa14/




Banta-Wright (2004) J Perinat Neonat Nurs **18**: 41

**General Population** → **MS/MS Newborn Screening** → **MS/MS positive**

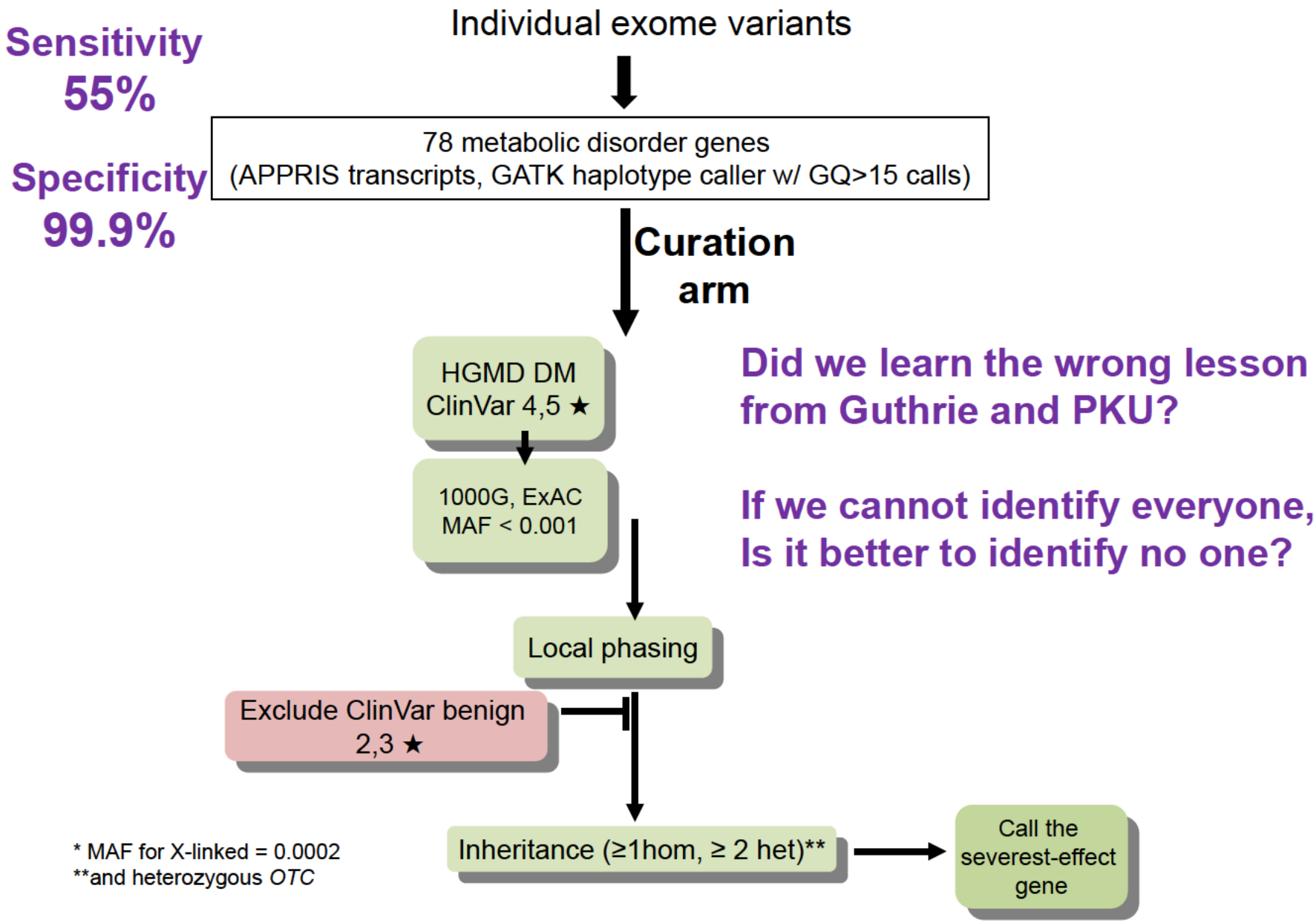
**Missed cases render exome sequencing alone unsuitable for population newborn screening of metabolic disorders**

**In several MS/MS screen positive cases, sequence data provided information that would help inform diagnosis**

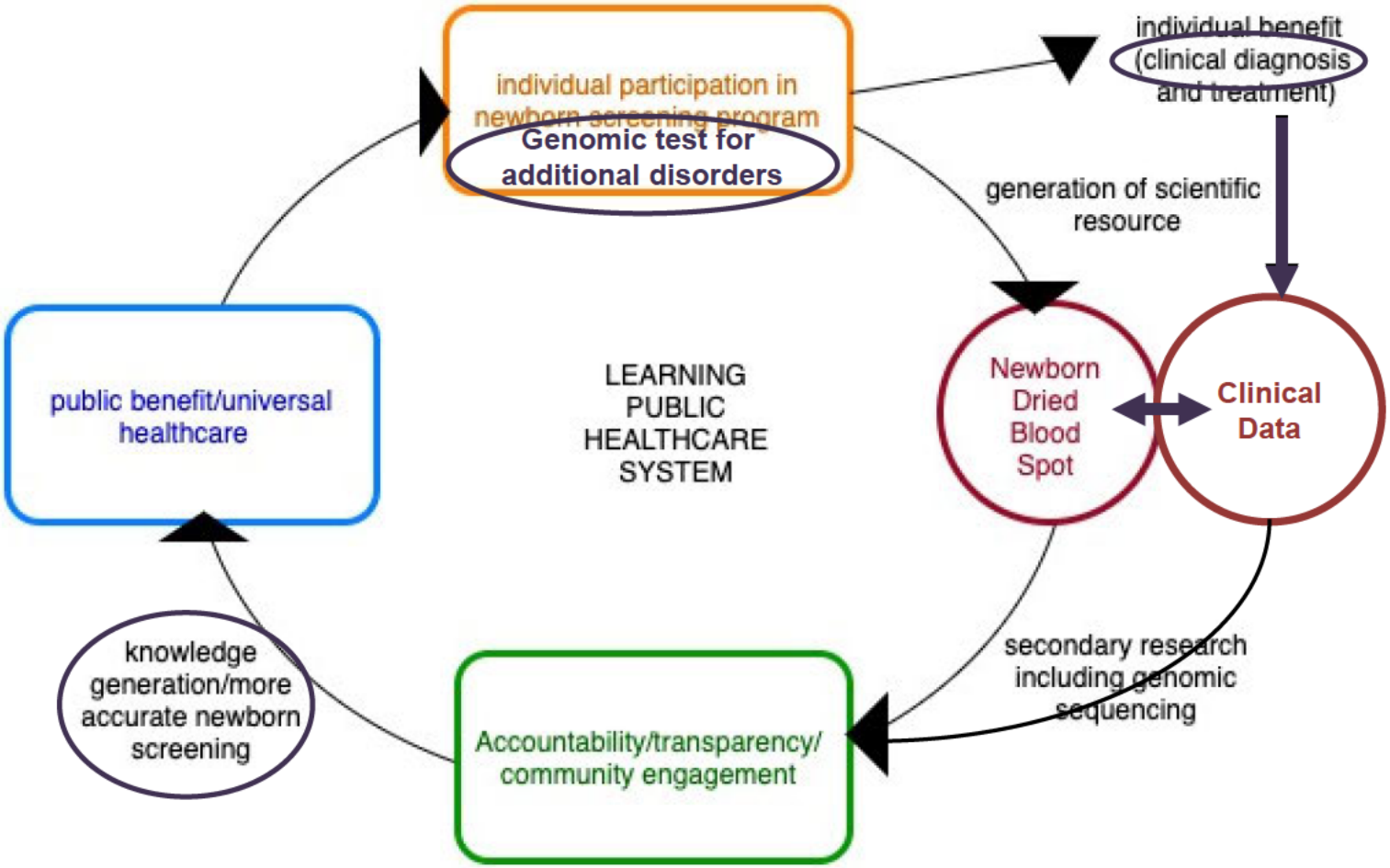
### Potential for sequencing exists in:

- NICU (high MS/MS false positives)
- Screening of specific disorders
- Expanding newborn screening for more genetic disorders
- Dissecting the allelic complexity in metabolic disorders
- Considering genomes, which reveal more than exomes
- Future improvements in technology and interpretation

# Variant interpretation pipeline



# An ethical and equitable Learning Public Health System




Based on work with Julia E.H. Brown, Matthew Norstad, Jodi Halpern, David C. Magnus, Barbara A. Koenig

# NBSeq


Aashish Adhikari


Yaqiong Wang

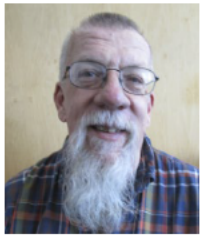

Yangyun Zou

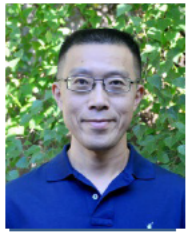

Flavia Chen


Jennifer Puck, PI


Pui Kwok, PI


Barbara Koenig, PI


Bob Currier


Hao Tang


Neil Risch


Joseph Sheih


Bob Nussbaum


Rajgopal Srinivasan


Uma Sunderam


Renata Gallagher


Sean Mooney

Beatriz Anguiano, Laia Bars, Amy Brower, Ajithavalli Chellappan, Binghuang Cai, Brad Dispensa, Carol Fraser-Browne, George Freedman, Julie Harris-Wai, Josie Johnston, Galen Joseph, Mark Kvale, Jamie King, Richard Lao, Dae Lee, Jeremy Michelson, Alan Nguyen, Erik Parens, Monica Smith, Dedeepya Vaka, Eunice Wan, Rachel Zacharias, Brandon Zerbe
Joy Boyer, Melissa Parisi, Tiina Urv, Anastasia Wise

UCSF University of California San Francisco

Berkeley UNIVERSITY OF CALIFORNIA

UW Medicine UW SCHOOL OF MEDICINE BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

TATA CONSULTANCY SERVICES

CONSORTIUM ON LAW, SCIENCE & HEALTH POLICY

NIH *Eunice Kennedy Shriver* National Institute of Child Health and Human Development

National Human Genome Research Institute

U19 HD077627