Leveraging Whole Genome Annotation for Genotype-Phenotype Association Studies

Eric Boerwinkle

ENCODE

Bethesda, 2015

UTHealth
The University of Texas
Health Science Center at Houston

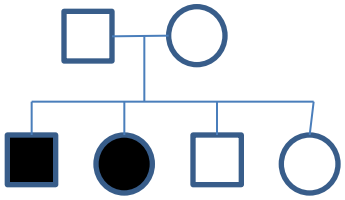
BCM
Baylor College of Medicine

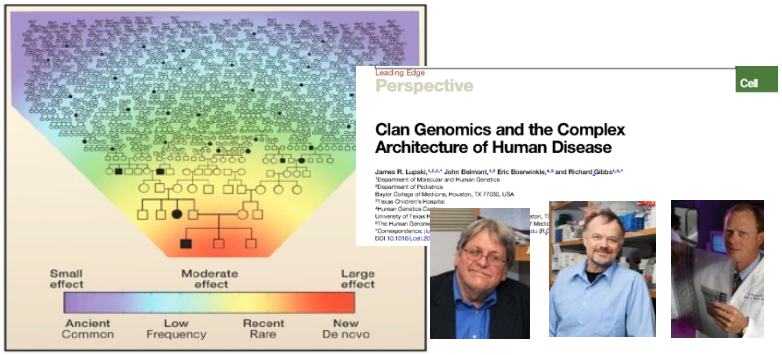
HGSC
HUMAN GENOME SEQUENCING CENTER

Clinical and Biologic Continuum of Simple/Complex and Rare/Common

Rare Mendelian…….. ‘Oligogenic’ ……. Common Complex

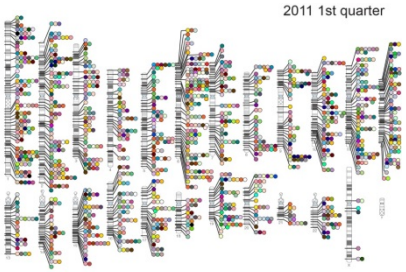
Low frequency/High impact
Not related to sporadic cases
in a simple way

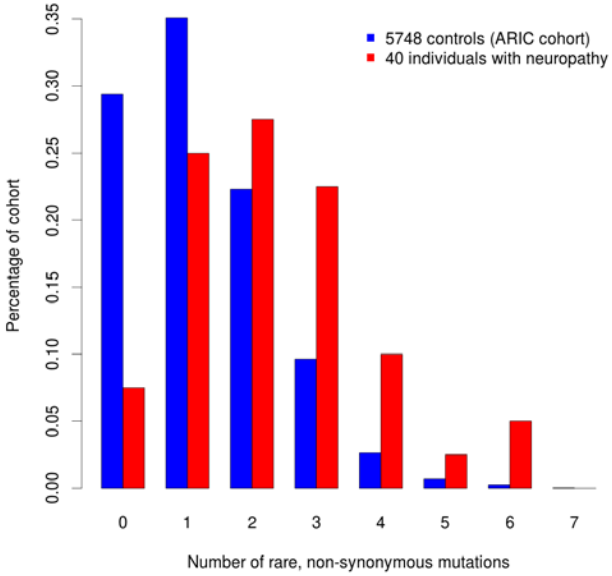


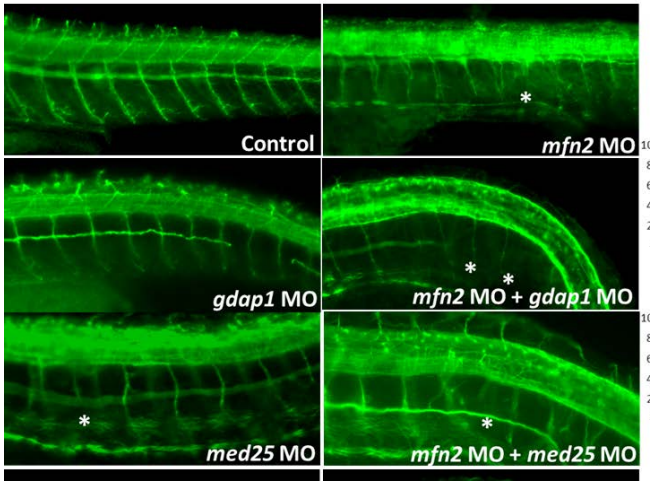


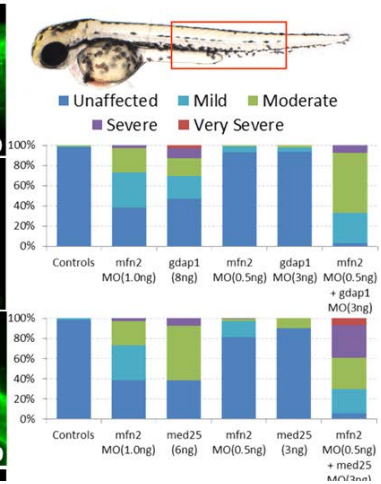
Jim Lupski Richard Gibbs Eric Boerwinkle

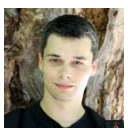
Few Actionable Alleles
Missing ‘Heritability’











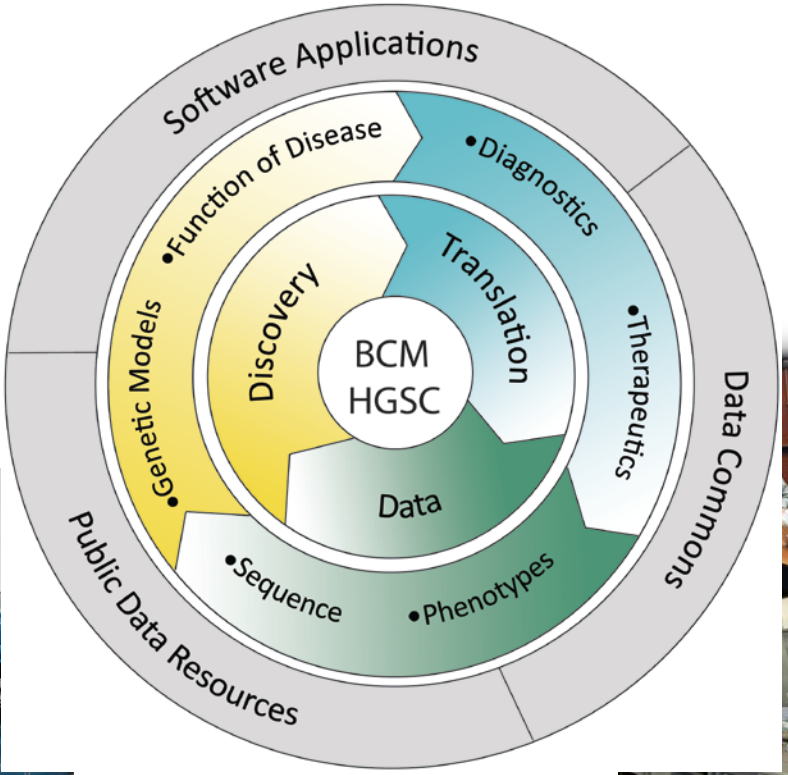
Clinical ⇔ Research Enterprises

The New York Times ON THE WEB

DNA Test for Rare Disorders Becomes More Routine,
G. Kolata

HGSC
HUMAN GENOME SEQUENCING CENTER
Baylor College of Medicine

WGL
WHOLE GENOME LABORATORY



Patients *versus* Research Subjects

Fire-Wall

HIPAA
Health Insurance Portability and Accountability Act

etc

- Privacy protected,
- Strict rules
- Generalized individual consent process
- Straightforward

- Privacy protected,
- Strict rules
- Customized consent process
- Complicated
- Includes data release/sharing options

Physician guided ⟷ ***Research protocol directed***

Clinical variant data sharing: technically easy
Restricted by HIPAA

Mendelian Discovery Scorecard

AJMG AMERICAN JOURNAL OF medical genetics

Research Article

Whole-exome sequencing identifies compound heterozygous mutations in *WDR62* in siblings with recurrent polymicrogyria†

David R. Murdock[1], Gary D. Clark[2,3], Matthew N. Bainbridge[1], Irene Newsham[1], Yuan-Qing Wu[1], Donna M. Muzny[1], Sau Wai Cheung[4], Richard A. Gibbs[1], Melissa B. Ramocki[2,3,*]

Article first published online: 10 AUG 2011
DOI: 10.1002/ajmg.a.34165
Copyright © 2011 Wiley-Liss, Inc.

Issue
American Journal of Medic Genetics Part A
Volume 155, Issue 9, page 2071–2077, September 20

Original Investigation

Mutations in *VRK1* Associated With Complex Motor and Sensory Axonal Neuropathy Plus Microcephaly

Claudia Gonzaga-Jauregui, PhD; Timothy Lotze, MD; Leila Jamal, ScM; Samantha Penney, MSc; Ian M. Campbell, BS; Davut Pehlivan, MD; Jill V. Hunter, MBBS; Suzanne L. Woodbury, MD; Gerald Raymond, MD; Adekunle M. Adesina, MD, PhD; Shalini N. Jhangiani, MS; Jeffrey G. Reid, PhD; Donna M. Muzny, MSc; Eric Boerwinkle, PhD; James R. Lupski, MD, PhD, DSc; Richard A. Gibbs, PhD; Wojciech Wiszniewski, MD, PhD

Genetics All rights reserved.
uary 2012

ling a Histone Acetyltransferase, Cause Genitopatellar Syndrome

Philippe M. Campeau[1, 17], Jaeseung C. Kim[2, 17], James T. Lu[3, 4], Jeremy A. Schwartzentruber[5], Omar A. Abdul-Rahman[6], Silke Schlaubitz[1], David M. Murdock[3], Ming-Ming Jiang[1], Edward J. Lammer[7], Gregory M. Enns[8], William J. Rhead[9], Jon Rowland[10], Stephen P. Robertson[11], Valérie Cormier-Daire[12], Matthew N. Bainbridge[3, 4], Xiang-Jiao Yang[13, 14], Marie-Claude Gingras[1, 3], Richard A. Gibbs[1, 3], David S. Rosenblatt[2, 14, 15], Jacek Majewski[2, 5] and Brendan H. Lee[1, 16]

[1] Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
[2] Department of Human Genetics, McGill University, Montreal, QC H3A 1B1, Canada

The NEW ENGLAND JOURNAL of MEDICINE

HOME | ARTICLES & MULTIMEDIA | ISSUES | SPECIALTIES & TOPICS | FOR AUTHORS | CME

ORIGINAL ARTICLE
BRIEF REPORT

WNT1 Mutations in Early-Onset Osteoporosis and Osteogenesis Imperfecta

...eau, M.D., Riku Kiviranta, M.D., Ph.D., Kati ...n, Ph.D., Maija Wessman, Ph.D., Terhi J. ...D., Ph.D., Heikki Kröger, M.D., Ph.D., William ...B.S., Deborah Krakow, M.D., Cynthia J.R. ...ee, M.D., Ph.D., and Outi Mäkitie, M.D., Ph.D.

...215458

Molecular Vision 2011; 17:3529-3540 <http://www.molvis.org/molvis/v17/a379>
Received 18 May 2011 | Accepted 26 December 2011 | Published 30 December 2011
© 2011 Molecular Vision

Exome capture sequencing identifies a novel mutation in *BBS4*

Hui Wang,[1,2] Xianfeng Chen,[2] Lynn Dudinsky,[9] Claire Patenia,[1] Yiyun Chen,[2] Yumei Li,[1,2] Yue Wei,[10] Emad B. Abboud,[11] Ali A. Al-Rajhi,[11] Richard Alan Lewis,[2,3,4] James R. Lupski,[2,4,5] Graeme Mardon,[2,5,6,7,8] Richard A. Gibbs,[1,2] Brian D. Perkins,[9] Rui Chen[1,2,8]

[1]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; [2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; [3]Department of Ophthalmology, Baylor College of Medicine, Houston, TX; [4]Texas Children's Hospital, Baylor College of Medicine, Houston, TX; [5]Department of Neurology, Baylor College of Medicine, Houston, TX; [6]Department of Neuroscience, Baylor College of Medicine, Houston, TX; [7]Department of Pathology, Baylor College of Medicine, Houston, TX; [8]Program in Developmental Biology, Baylor College of Medicine, Houston, TX; [9]Department of Biology, Texas A&M University, College Station, TX; [10]Leukemia Department, University of Texas, M. D. Anderson Cancer Center, Houston, TX; [11]King Khaled Eye Specialist Hospital, Riyadh, Kingdom of Saudi Arabia

Journal of Thrombosis and Haemostasis, **11**: 1228–1239
DOI: 10.1111/jth.12291

ORIGINAL ARTICLE

Next-generation sequencing study finds an excess of rare, coding single-nucleotide variants of *ADAMTS13* in patients with deep vein thrombosis

L. A. LOTTA,*† G. TUANA,*‡ J. YU,† I. MARTINELLI,* M. WANG,† F. YU,† S. M. PASSAMONTI,* E. PAPPALARDO,* C. VALSECCHI,* S. E. SCHERER,† W. HALE IV,† D. M. MUZNY,† G. RANDI,§ F. R. ROSENDAAL,*¶ R. A. GIBBS†[1] and F. PEYVANDI*[1]
*Angelo Bianchi Bonomi Hemophilia and Thrombosis Center, Fondazione IRCCS Ca' Granda – Ospedale Maggiore Policlinico, Università degli Studi di Milano and Fondazione Luigi Villa, Milan, Italy; †Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; ‡Hematology 1 CTMO, Fondazione IRCCS Ca' Granda – Ospedale Maggiore Policlinico; §Department of Clinical Sciences and Community Health, Università degli Studi di Milano, Milan, Italy; and ¶Departments of Clinical Epidemiology, Thrombosis and Hemostasis, Leiden University Medical Center, Leiden, the Netherlands

Journal of Bone and Mineral Research
JBMR

JOURNAL INFORMATION
- Home
- Aims and Scope
- Editorial Board
- News
- Accepted Articles

Home > Current Issue > Article

Mutations in *SERPINF1* cause Osteogenesis imperfecta Type VI

Erica P. Homan, Frank Rauch, Ingo Grafe, Caressa Lietman, Jennifer A. Doll, Brian Dawson, Terry Bertin, Dobrawa Napierala, Roy Morello, Richard Gibbs, Lisa White, Rika

Human Mutation
Variation, Informatics, and Disease

Research Article

Whole-exome sequencing identifies *ALMS1*, *IQCB1*, *CNGA3*, and *MYO7A* mutations in patients with leber congenital amaurosis†

Xia Wang[1,2], Hui Wang[1,2,2], Ming Cao[2], Zhe Li[3], Xianfeng Chen[1], Claire Patenia[2], Athurva Gore[3], Emad B. Abboud[4], Ali A. Al-Rajhi[4], Richard A. Lewis[1,5,6], James R. Lupski[1,6,7], Graeme Mardon[1,7,8,9,10], Kun Zhang[3], Donna Muzny[1,2], Richard A. Gibbs[1,2], Rui Chen[1,2,10,*]

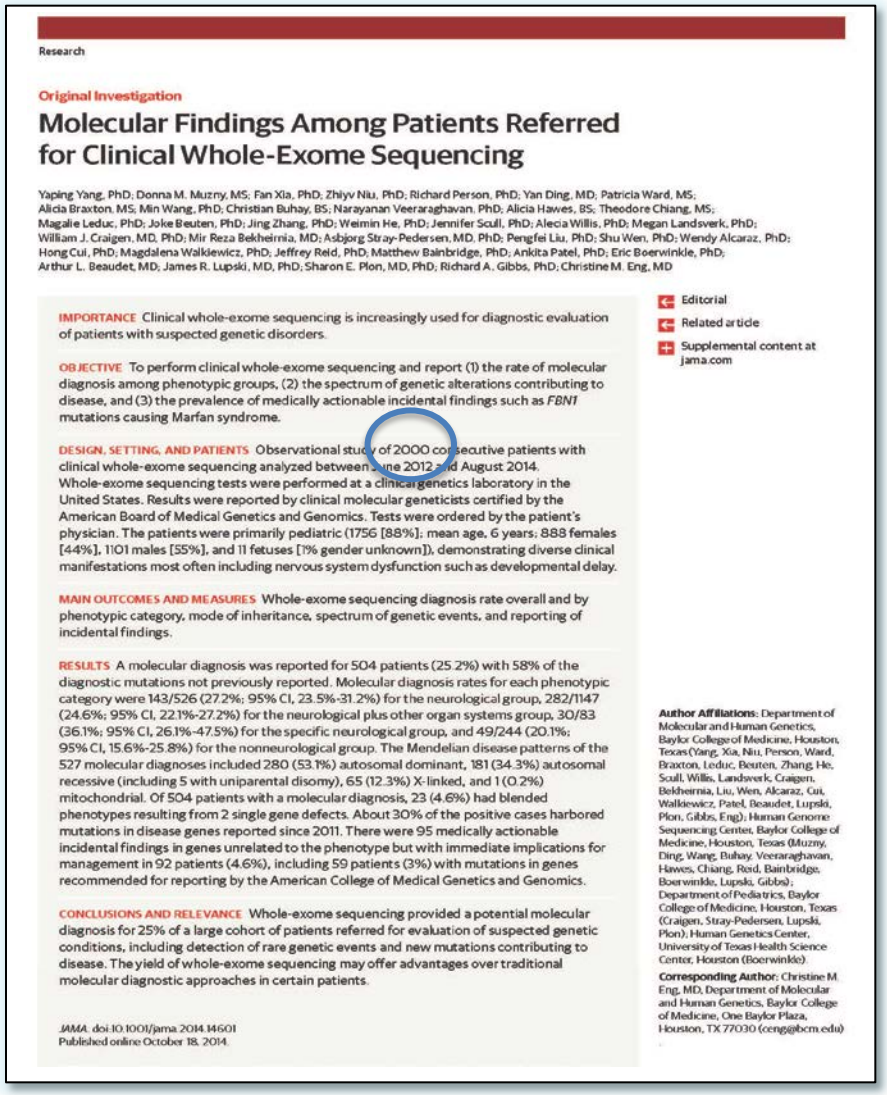
Article first published online: 23 SEP 2011
DOI: 10.1002/humu.21587

Issue
Human Mutation
Early View (Online Version of Record published before inclusion in an issue)

REPORT

A Recurrent *PDGFRB* Mutation Causes Familial Infantile Myofibromatosis

Yee Him Cheung,[1,20] Tenzin Gayden,[2,20] Philippe M. Campeau,[3] Charles A. LeDuc,[1] Donna Russo,[1] Van-Hung Nguyen,[4] Jiancheng Guo,[1] Ming Qi,[5,6,7] Yanfang Guan,[7] Steffen Albrecht,[4] Brenda Moroz,[8] Karen W. Eldin,[9] James T. Lu,[10,11] Jeremy Schwartzentruber,[12] David Malkin,[13] Albert M. Berghuis,[14] Sherif Emil,[15] Richard A. Gibbs,[10] David L. Burk,[14] Megan Vanstone,[16] Brendan H. Lee,[3,17] David Orchard,[18] Kym M. Boycott,[16] Wendy K. Chung,[1,*] and Nada Jabado[2,19,*]



Research

Original Investigation

Molecular Findings Among Patients Referred for Clinical Whole-Exome Sequencing

Yaping Yang, PhD; Donna M. Muzny, MS; Fan Xia, PhD; Zhiyv Niu, PhD; Richard Person, PhD; Yan Ding, MD; Patricia Ward, MS; Alicia Braxton, MS; Min Wang, PhD; Christian Buhay, BS; Narayanan Veeraraghavan, PhD; Alicia Hawes, BS; Theodore Chiang, MS; Magalie Leduc, PhD; Joke Beuten, PhD; Jing Zhang, PhD; Weimin He, PhD; Jennifer Scull, PhD; Alecia Willis, PhD; Megan Landsverk, PhD; William J. Craigen, MD, PhD; Mir Reza Bekheirnia, MD; Asbjorg Stray-Pedersen, MD, PhD; Pengfei Liu, PhD; Shu Wen, PhD; Wendy Alcaraz, PhD; Hong Cui, PhD; Magdalena Walkiewicz, PhD; Jeffrey Reid, PhD; Matthew Bainbridge, PhD; Ankita Patel, PhD; Eric Boerwinkle, PhD; Arthur L. Beaudet, MD; James R. Lupski, MD, PhD; Sharon E. Plon, MD, PhD; Richard A. Gibbs, PhD; Christine M. Eng, MD

IMPORTANCE Clinical whole-exome sequencing is increasingly used for diagnostic evaluation of patients with suspected genetic disorders.

OBJECTIVE To perform clinical whole-exome sequencing and report (1) the rate of molecular diagnosis among phenotypic groups, (2) the spectrum of genetic alterations contributing to disease, and (3) the prevalence of medically actionable incidental findings such as *FBN1* mutations causing Marfan syndrome.

DESIGN, SETTING, AND PATIENTS Observational study of 2000 consecutive patients with clinical whole-exome sequencing analyzed between June 2012 and August 2014. Whole-exome sequencing tests were performed at a clinical genetics laboratory in the United States. Results were reported by clinical molecular geneticists certified by the American Board of Medical Genetics and Genomics. Tests were ordered by the patient's physician. The patients were primarily pediatric (1756 [88%]; mean age, 6 years; 888 females [44%], 1101 males [55%], and 11 fetuses [1% gender unknown]), demonstrating diverse clinical manifestations most often including nervous system dysfunction such as developmental delay.

MAIN OUTCOMES AND MEASURES Whole-exome sequencing diagnosis rate overall and by phenotypic category, mode of inheritance, spectrum of genetic events, and reporting of incidental findings.

RESULTS A molecular diagnosis was reported for 504 patients (25.2%) with 58% of the diagnostic mutations not previously reported. Molecular diagnosis rates for each phenotypic category were 143/526 (27.2%; 95% CI, 23.5%-31.2%) for the neurological group, 282/1147 (24.6%; 95% CI, 22.1%-27.2%) for the neurological plus other organ systems group, 30/83 (36.1%; 95% CI, 26.1%-47.5%) for the specific neurological group, and 49/244 (20.1%; 95% CI, 15.6%-25.8%) for the nonneurological group. The Mendelian disease patterns of the 527 molecular diagnoses included 280 (53.1%) autosomal dominant, 181 (34.3%) autosomal recessive (including 5 with uniparental disomy), 65 (12.3%) X-linked, and 1 (0.2%) mitochondrial. Of 504 patients with a molecular diagnosis, 23 (4.6%) had blended phenotypes resulting from 2 single gene defects. About 30% of the positive cases harbored mutations in disease genes reported since 2011. There were 95 medically actionable incidental findings in genes unrelated to the phenotype but with immediate implications for management in 92 patients (4.6%), including 59 patients (3%) with mutations in genes recommended for reporting by the American College of Medical Genetics and Genomics.

CONCLUSIONS AND RELEVANCE Whole-exome sequencing provided a potential molecular diagnosis for 25% of a large cohort of patients referred for evaluation of suspected genetic conditions, including detection of rare genetic events and new mutations contributing to disease. The yield of whole-exome sequencing may offer advantages over traditional molecular diagnostic approaches in certain patients.

Editorial

Related article

Supplemental content at jama.com

Author Affiliations: Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas (Yang, Xia, Niu, Person, Ward, Braxton, Leduc, Beuten, Zhang, He, Scull, Willis, Landsverk, Craigen, Bekheirnia, Liu, Wen, Alcaraz, Cui, Walkiewicz, Patel, Beaudet, Lupski, Plon, Gibbs, Eng); Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas (Muzny, Ding, Wang, Buhay, Veeraraghavan, Hawes, Chiang, Reid, Bainbridge, Boerwinkle, Lupski, Gibbs); Department of Pediatrics, Baylor College of Medicine, Houston, Texas (Craigen, Stray-Pedersen, Lupski, Plon); Human Genetics Center, University of Texas Health Science Center, Houston (Boerwinkle).

Corresponding Author: Christine M. Eng, MD, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030 (ceng@bcm.edu).

JAMA. doi:10.1001/jama.2014.14601
Published online October 18, 2014.

JAMA®

The Journal of the American Medical Association

Y Yang and coauthors

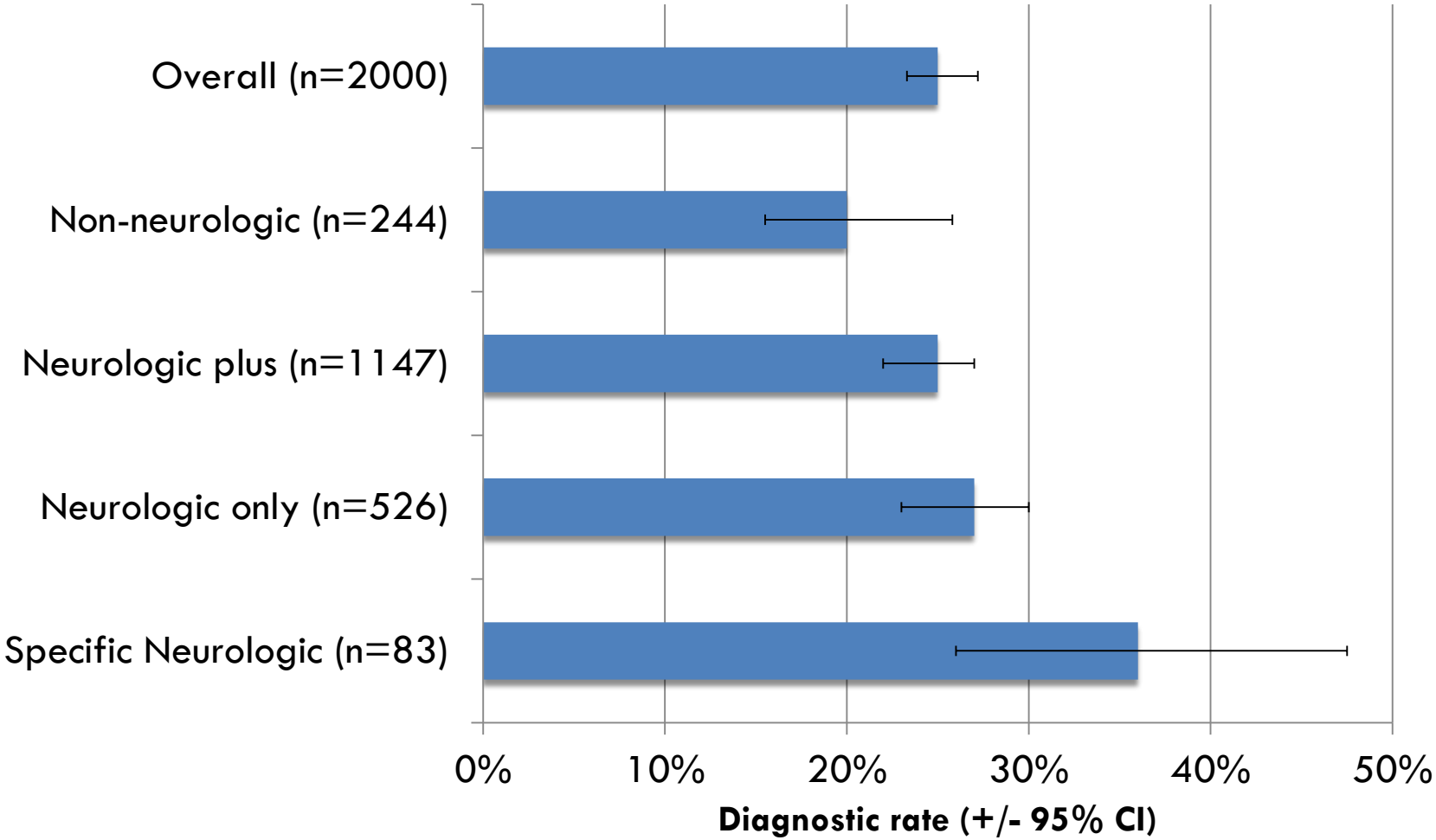
Molecular Findings Among Patients Referred for Clinical Whole-Exome Sequencing

Published online October 18, 2014

Available at jama.com and on The JAMA Network Reader at mobile.jamanetwork.com

The JAMA Network

Molecular Diagnosis Rate

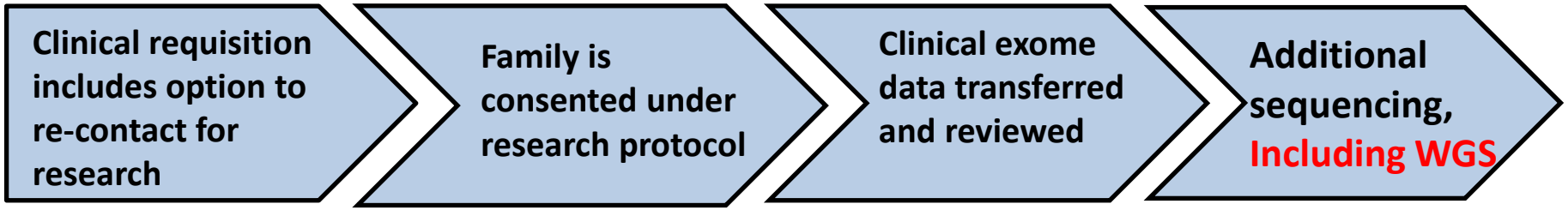


25% Diagnostic Rate Maintained for 3384 patients

Yang, et al., *JAMA,* 2014

Research Pipeline for Unsolved Cases

- Unsolved cases available for research consent through the CMG: 1,697
- Cases consented for research: 121
- Trios completed exome sequencing: 69



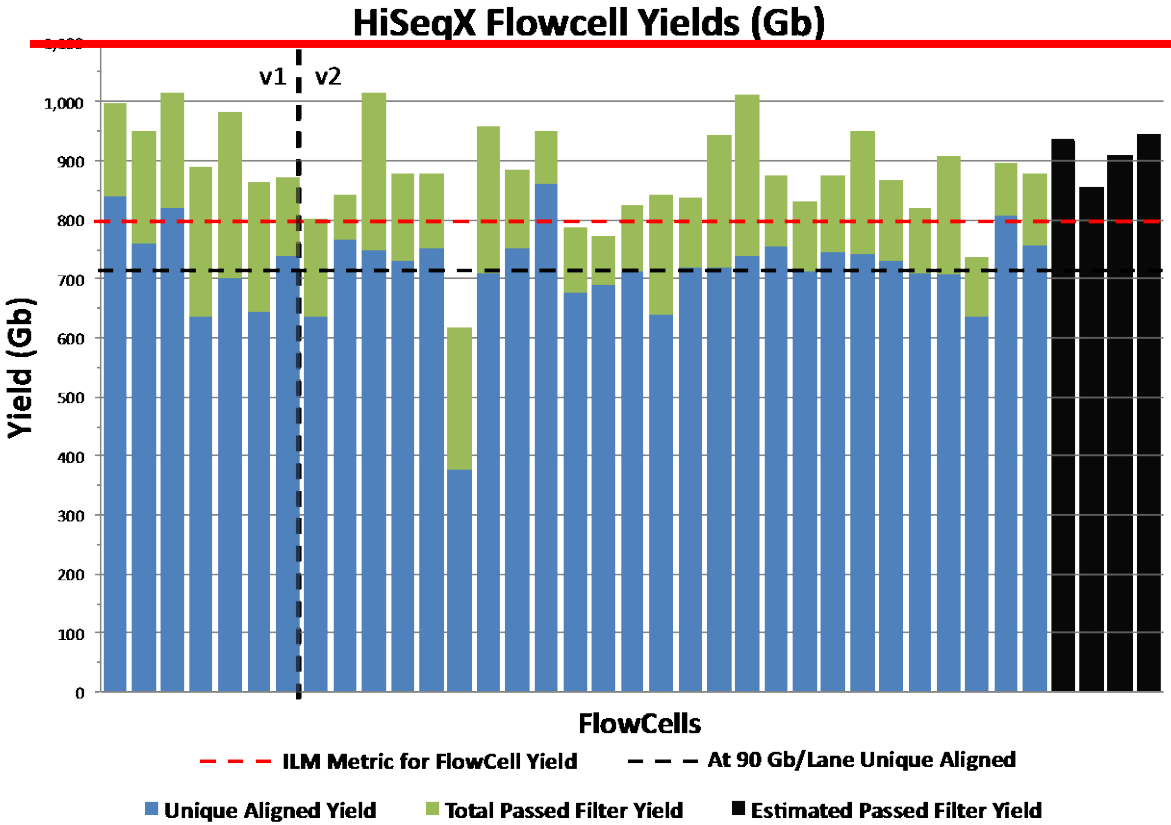
Phenotype	Gene	Classification	# Cases	Notes
DD, Lt Vt NonCompaction, Abnormal muscle biopsy	MIPEP	Novel	1	
DD, Epilepsy, Ataxia, Hypotonia, Hirschsprung disease	PMPCA	Novel	3*	Validation in Progress
Abnormal gait, slurred speech, myopathic facies, hypothyroidism	TANGO2	Novel	2*	
Neonatal Seizures	SLC13A5	Known	2*	Clinical Re-Signout
Encephalopathy, dysmorphic features, bilateral hearing loss hypopigmentation, agenesis corpus callosum, behavioral problems	ZBTB20	Known	1	Clinical Re-Signout
NK cell deficiency, GH treated Short Stature	PIK3CD	Known	3*	Validation in Progress

**Additional Cases from BHCMG database*

HiSeq XTen Validation and Burn-in

- First instrument: Sept 2014
- Validation: Hapmap trio + HS1011 (CMT neuropathy sample)
- Burn-in: targeting 500 production samples
  - ARIC, Mendelian, Cancer cell lines, etc
  - Version 1 and Version 2 chemistry and software
  - Standard Nano libraries and PCR free libraries
  - Single samples per lane and multiplexing
- Final goal: implementation for clinical WGS sequencing

Whole Genome Sequencing: XTen



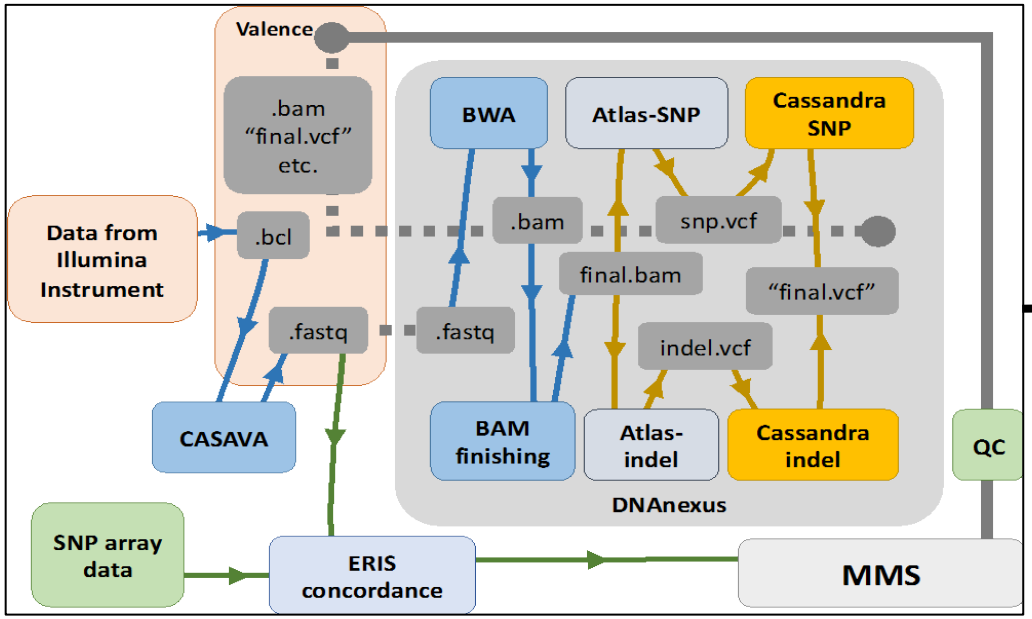
Total samples = ~300
Avg Pass filter = 881 Gb
Avg Unique Aligned = 717 Gb

Variant Concordance Analysis

Sample A	Average Coverage	Sensitivity	False Positive Rate
HS1011 Nano_v1	36X	96.79%	2.77%
HS1011 Nano_v2	36X	97.43%	3.80%
HS1011 HiseqX PCR free	35X	96.57%	3.28%

"Truth Set" is a PASS concordance of 4 WGS HS1011 samples

Tools Upgraded for WGS Analysis

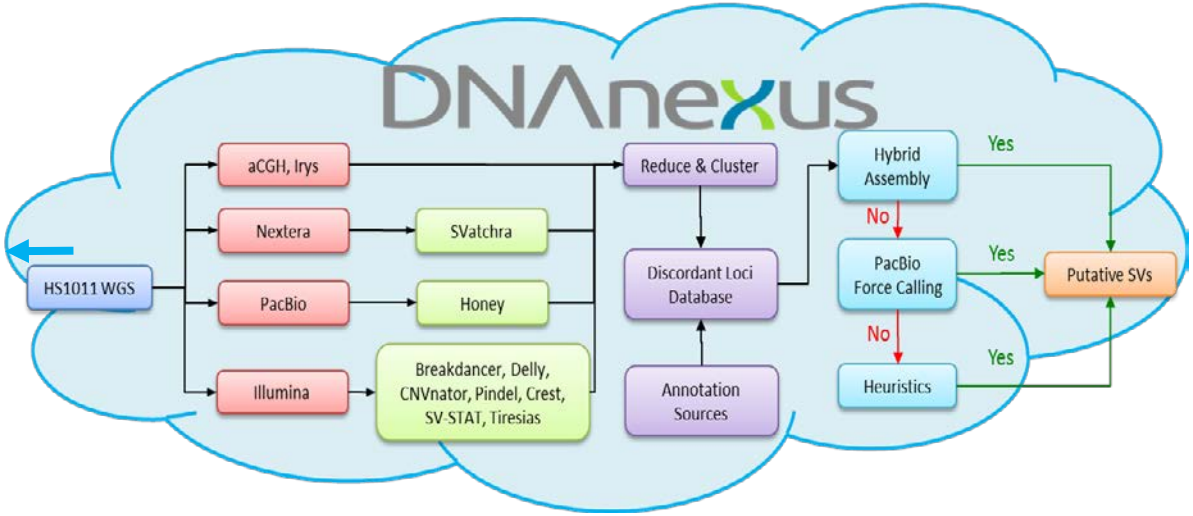


MercuryV2:

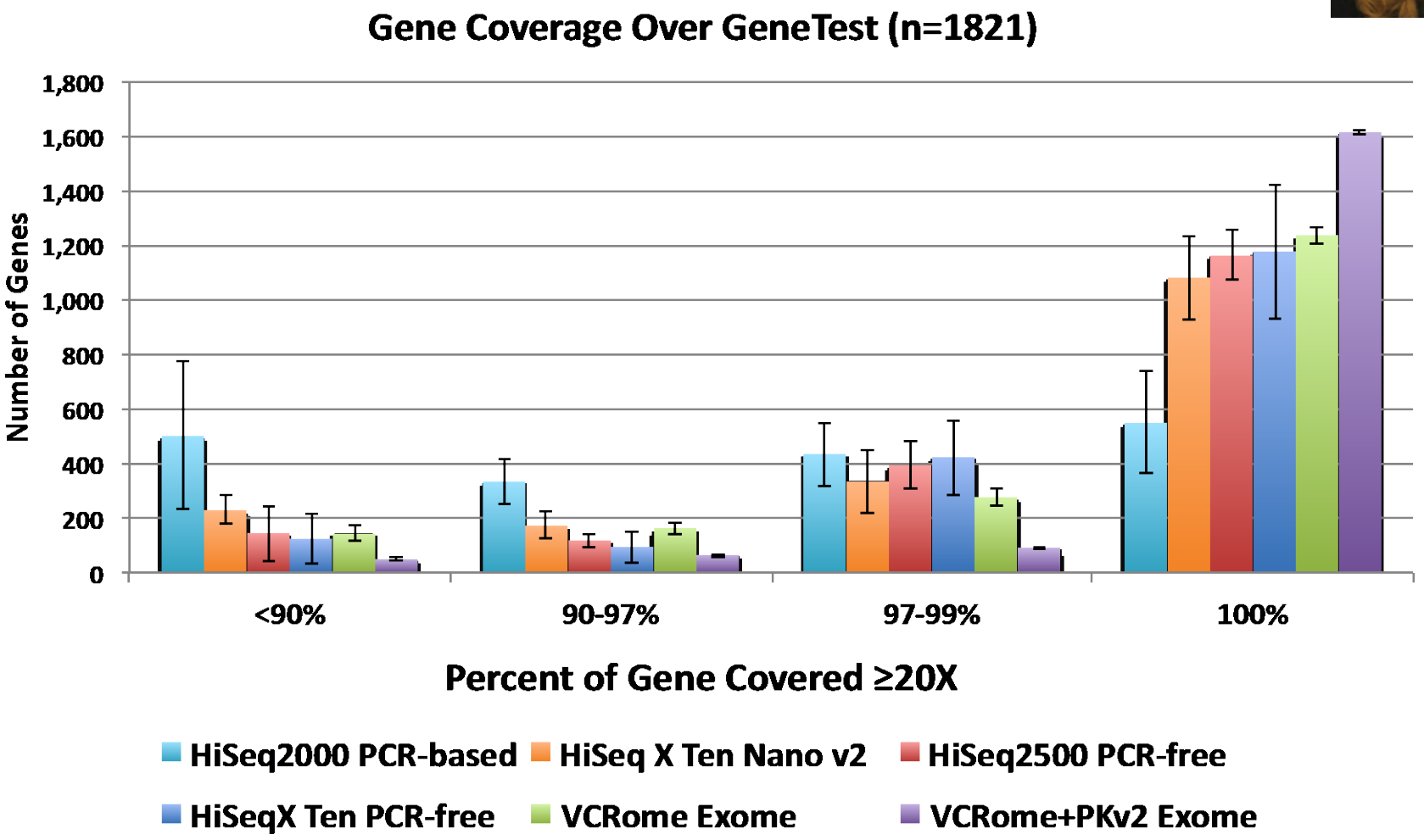
- Includes CasandraV2
- In-House developed annotation pipeline
- Upgrade includes ENCODE, FANTOM5, Ensembl, along with CADD, Funseq, RegulomeDB for scores to aid in prioritization

SV Parliament:

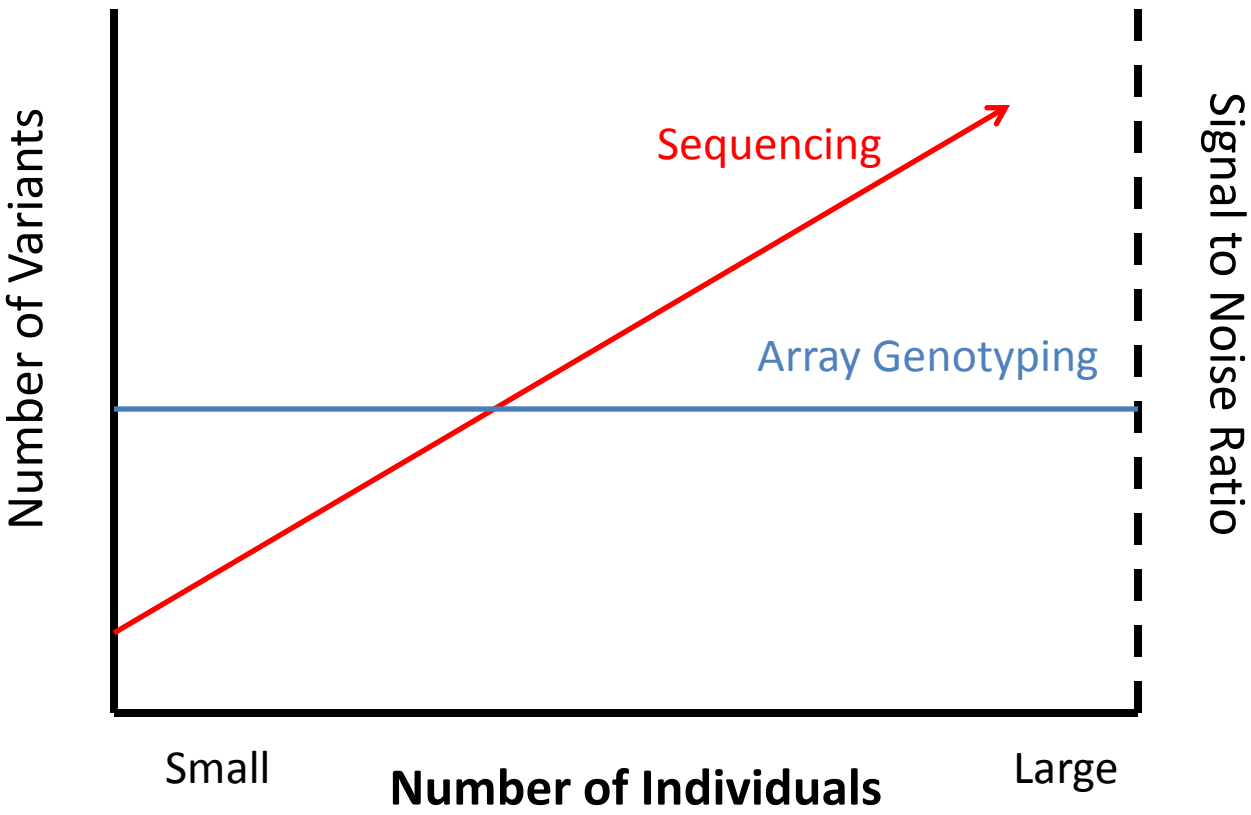
- Infrastructure for consensus structural variant calling using platform specific tools
- Technology-specific method to reduce clustering
- Local assembly for evaluation



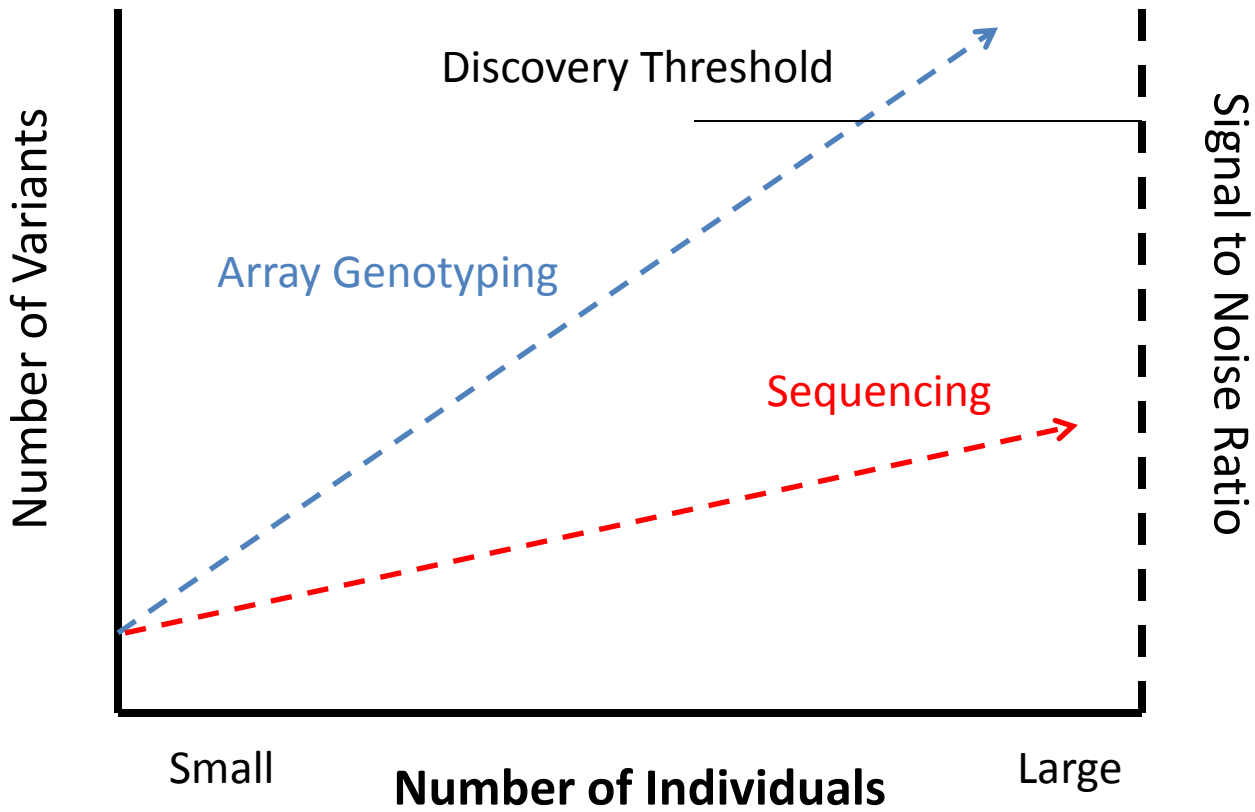
Genetest Coverage Comparison: 30X Genome and Clinical Exomes



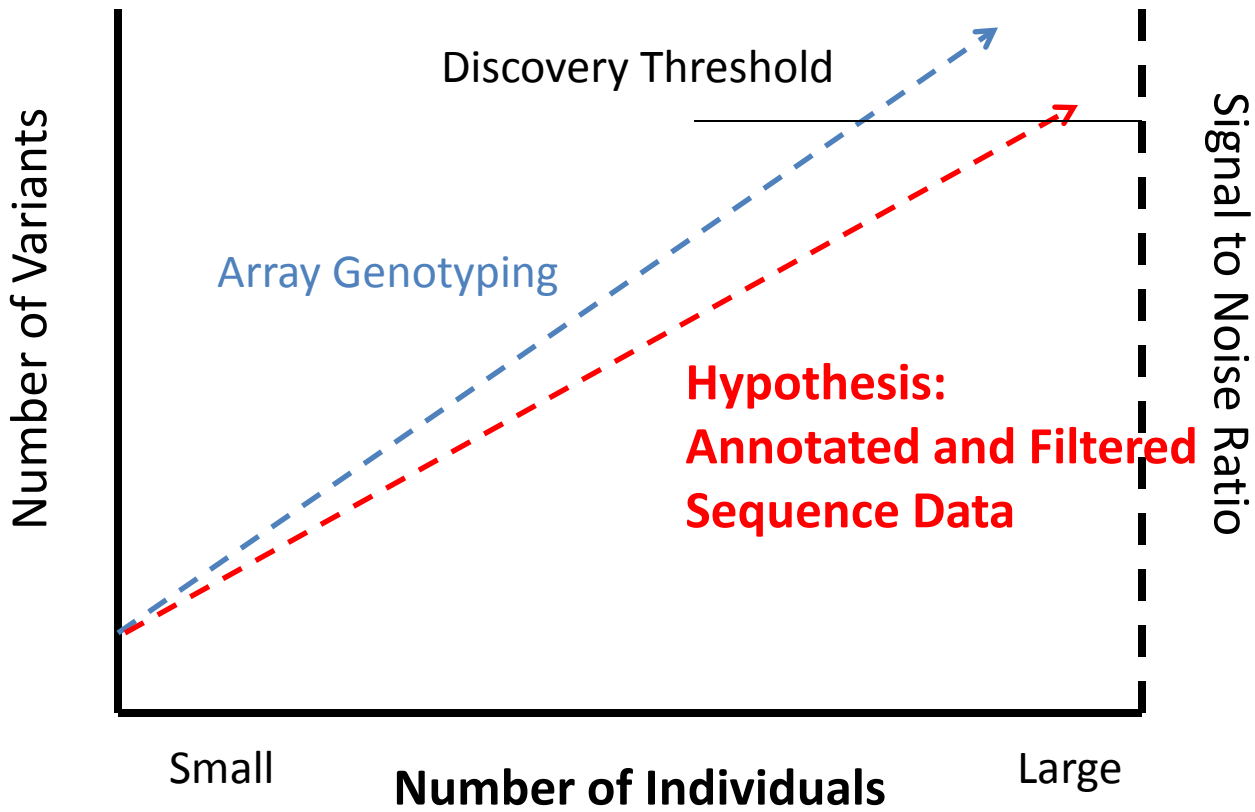
A Signal to Noise Problem



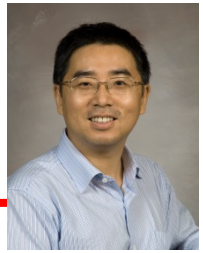
Signal to Noise Ratio

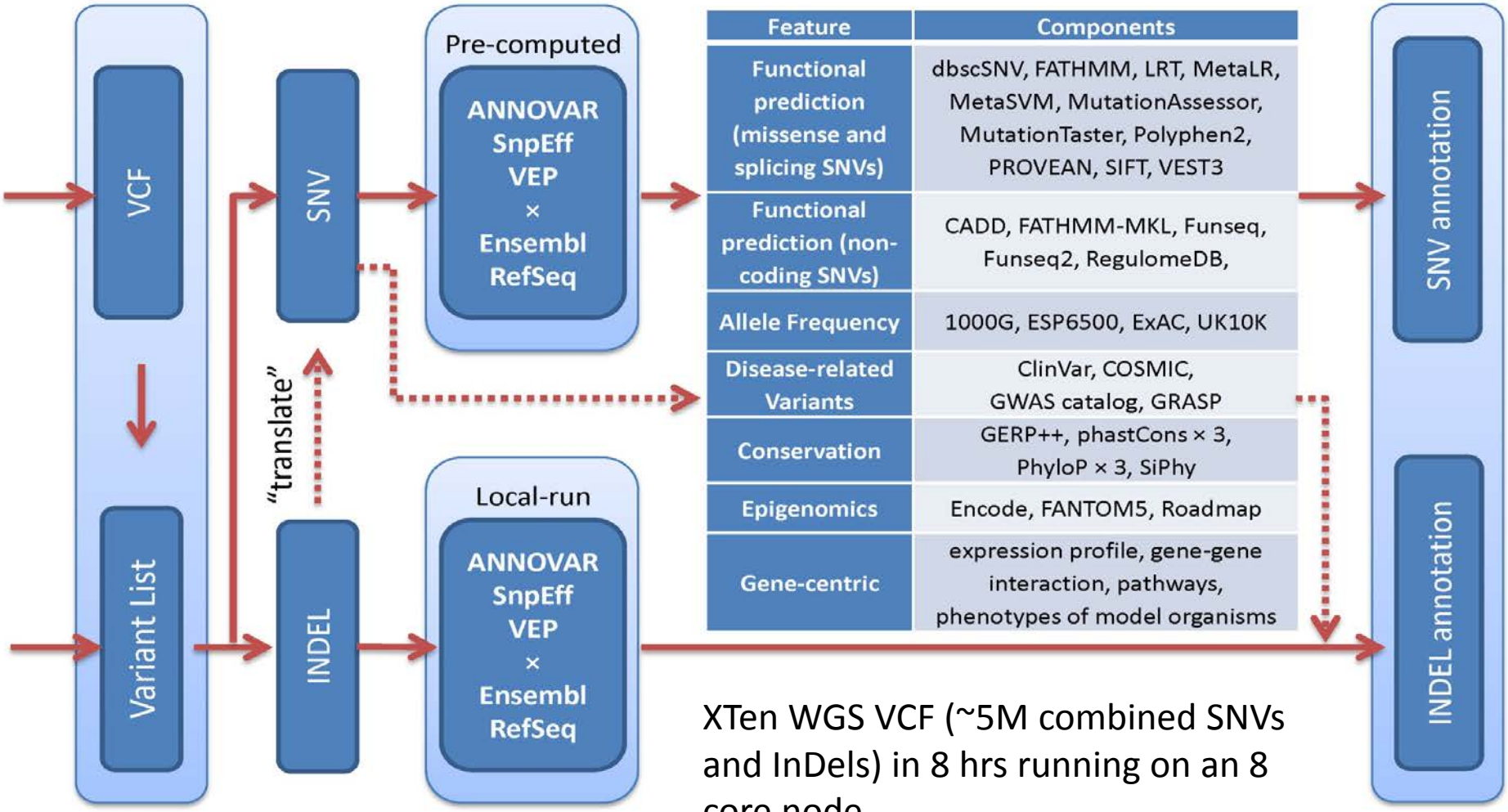


Signal to Noise Ratio



Cassandra XTen Annotation





Feature	Components
Functional prediction (missense and splicing SNVs)	dbscSNV, FATHMM, LRT, MetaLR, MetaSVM, MutationAssessor, MutationTaster, Polyphen2, PROVEAN, SIFT, VEST3
Functional prediction (non-coding SNVs)	CADD, FATHMM-MKL, Funseq, Funseq2, RegulomeDB,
Allele Frequency	1000G, ESP6500, ExAC, UK10K
Disease-related Variants	ClinVar, COSMIC, GWAS catalog, GRASP
Conservation	GERP++, phastCons × 3, PhyloP × 3, SiPhy
Epigenomics	Encode, FANTOM5, Roadmap
Gene-centric	expression profile, gene-gene interaction, pathways, phenotypes of model organisms

XTen WGS VCF (~5M combined SNVs and InDels) in 8 hrs running on an 8 core node

Pipeline for unsolved clinical cases 2: whole genome sequencing – HiSeq XTen

Overall Summary

Families	Samples	Family Structure	Phenotypes
10	31	7 Trios, 2 Quads & Discordant Twins	Cardiac, Neurologic, Immunodeficiencies and Renal

Average Quality Per Sample

Uniq. Aligned Gb	Coverage	Mode Insert Size	% Align	% Error	Unique %
111	37X	343	92	1.2	87

Strong Candidate Findings:

review by physician and/or validation in progress

Phenotype	Gene	Classification	# Cases
dysplastic R kidney, hypotonia, neurogenic bladder, nonverbal, motor regression	*EBF3*	Novel	3*
NBIA like (Neurodegeneration with Brain Iron Accumulation)	*OPA1*	Phenotypic Expansion?	2*
	190bp deletion in SP8	*additional analysis in progress*	

****Additional Cases found in BHCMG Database***

Family 9: Neurodegeneration with Brain Iron Accumulation

- Discovered a large (168bp) *de novo* deletion in the gene SP8 (not picked up by multiple tools in WES).
- The protein encoded by this gene is an SP family transcription factor that in mouse has been shown to be essential for proper limb development
- Although the phenotype isn't concordant with the patient, the importance of the gene and the deleteriousness of the variant warrant consideration

CHARGE –omics Ongoing Work

Ongoing/projected whole-genome omics data in CHARGE							
Cohort	**WES**	**WGS**	**Express**	**Methyl**	**Metab**	**miRNA**	**Flow Cy**
ARIC EA	10,000	1,706	2,000	1,000	1,616	2,000	1,498
ARIC AA	3,200	1,901	40	2,873	2,594	0	396
CHS	3,048	830	0	320	0	0	1,200
FHS	2,975	4,859	5,622	2,726	2,650	5,718	0
RS	4,500	100	1,781	1,700	2,000	0	0
CARDIA	204	0	0	0	1,000	0	0
FamHS	0	0	0	0	0	0	0
HABC	0	0	0	0	319	0	0
JHS AA	3,237	3,550	0	96	0	0	0
MESA	404	0	1,264	1,264	4,000	1,264	1,200
ERF	1569	25	0	0	3,094	0	0
Total	29,137	12,971	10,707	9,979	17,273	8,982	4,294

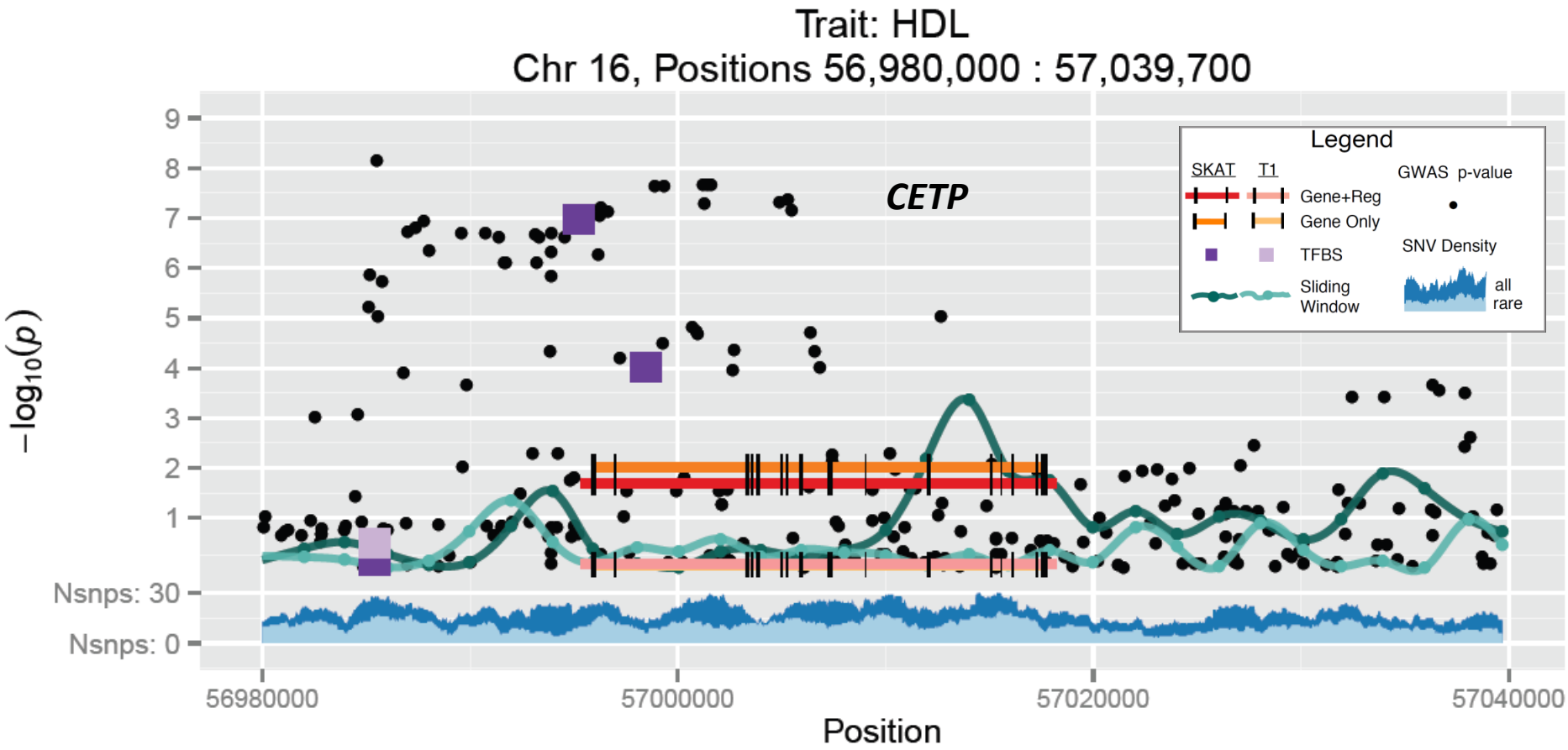
WES = whole exome sequencing; WGS = whole genome sequencing; Express = genome wide gene expression (array or sequencing); Methyl = whole genome methylation data; Metab = metabolomics data; miRNA = micro RNA qPCR, T-Cell = color flow cytometry panel

gSNAP: genomic Sequencing Nucleotide polymorphism Analysis Pipeline

Features:

- a large scale genomic computing framework, exploiting Cloud, superclusters and local HPC
- high quality *joint* variant calling at 10k-100k genomes scale
- one-stop shopping experience (multiple variant types with multiple callers)
  - SNP calling: GATK (HaplotypeCaller, UnifiedGenotyper), SNPTools, GotCloud
  - INDEL calling: Platypus, GATK, Pindel
- robust consensus strategy to resolve complexity in calling algorithms
- various filters to resolve complexities in genomic regions, variant types, and haplotype substructures
- highly scalable with inexpensive cost per genome
- applied to CHARGE WGS variant calling of 5300 samples

Lachesis: Full Model Annotation



Thank you and Questions?

Virtuous Cycle
- Gibbs

Clan Genomic
- Gibbs and Lupski

BMGL
- Eng, Yang, Muzny, Gibbs, Beudet

BHCMG
- Valle, Lupski, Jhangiani

CHARGE
- Psaty, O'Donnell, Hofman

CHARGE WGS
- Gibbs, Muzny, Morrison, Voorman

Cassandra
- White, Bainbridge, Liu

UTHealth™ The University of Texas Health Science Center at Houston

BCM Baylor College of Medicine

HGSC HUMAN GENOME SEQUENCING CENTER