Clinical genetics of neurodevelopmental disorders

Gholson J. Lyon, M.D. Ph.D.

for

CME accredited Seminar Series - Mount Sinai Medical Center.

Cold Spring Harbor Laboratory
Stanley Institute for Cognitive Genomics

Acknowledgments

Martin Reese
Edward Kiruluta

Barry Moore
Alan Rope
Jeffrey J Swensen
Lynn Jorde
Mark Yandell

Thomas Arnesen
Nathalie Reuter
Line Myklebust

Reid Robison

Jason O'Rawe
Yiyang Wu
Han Fang
Max Doerfel
Michael Schatz
Giuseppe Narzisi

our study families and many others

David Mittelman

Kai Wang

Tina Hambuch
Erica Davis
Dawn Barry

Ghent, Belgium
Petra Van Damme
Kris Gevaert

Jason O’Rawe

Yiyang Wu

Han Fang

Max Doerfel

Learning Objectives

- The participants will be able to discuss the extraordinary amount of variable expressivity seen in neurodevelopmental disorders.

- The participants will plan ways in which to integrate genomic and phenotypic longitudinal data to prevent the development of certain illnesses.

Conflicts of Interest

Advisory Boards

"Using a patented process invented by Dr. Lee Silver of Princeton University, we predict what would happen if the genes of two biological parents were to come together—well before there's a pregnancy. This technology allows us to analyze thousands of potential genetic combinations for millions of mutations that could affect the health of your future family".

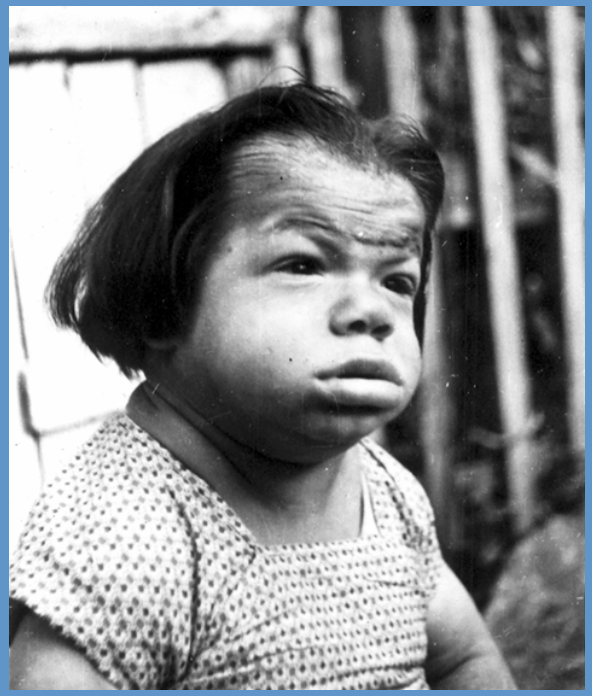
Learning Objectives

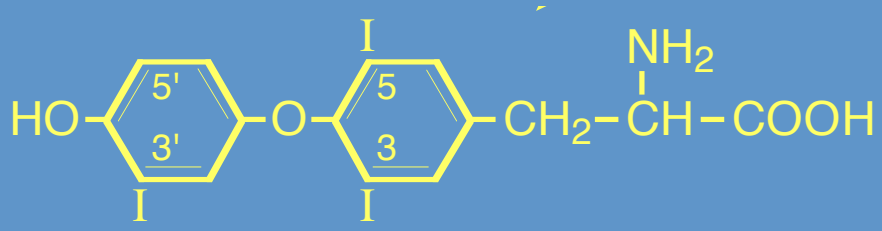
- **The participants will be able to discuss the extraordinary amount of variable expressivity seen in neurodevelopmental disorders.**

- The participants will plan ways in which to integrate genomic and phenotypic longitudinal data to prevent the development of certain illnesses.

The Story began for me at least by 1993….

when I joined the lab of Don St.Germain to study the role of thyroid hormone in cretinism, which is caused by lack of iodine during maternal pregnancy, so this is an environmentally triggered disease.

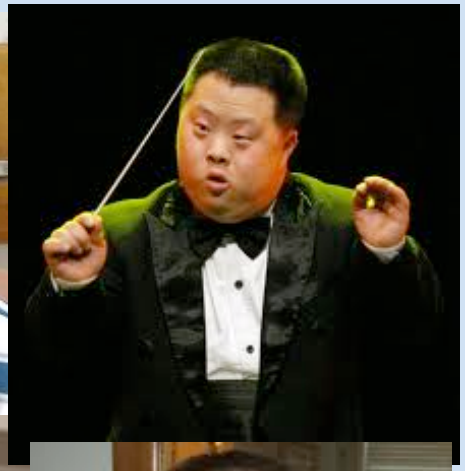
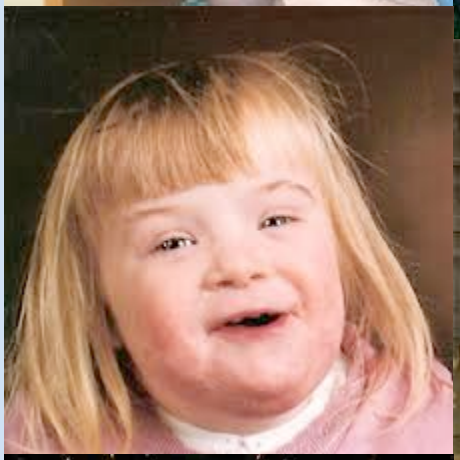




Thyroid Hormone

Down Syndrome

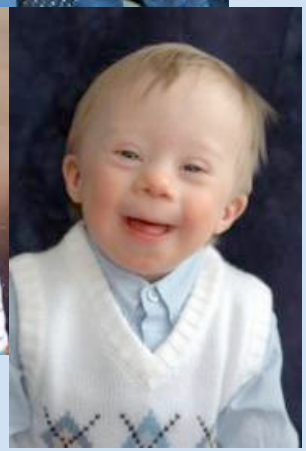










Hum Genet (2009) 126:317–324
DOI 10.1007/s00439-009-0690-1

HISTORICAL AND PERSONAL PERSPECTIVES

Fiftieth anniversary of trisomy 21: returning to a discovery

Marthe Gautier · Peter S. Harper

Published online: 30 June 2009

"In reality, discoveries are due to people at the edge of the formalised groups of researchers"

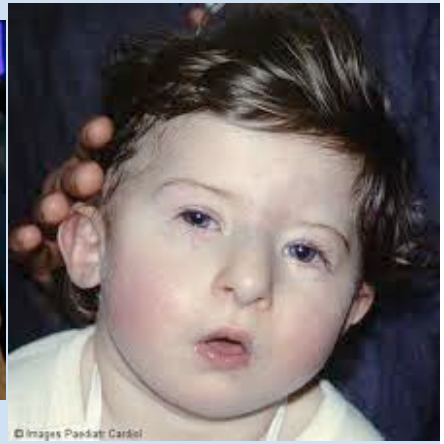
Pierre Laszlo

Fifty years ago, I was the co-author[1] of the first paper that showed the presence of an additional chromosome (Lejeune et al. 1959) in the syndrome identified by Langdon Down in 1866 and commonly known as "mongolism" in France at the time. This, the first autosomal chromosome aberration recognised in the cells of the human species, was named trisomy 21. I thought it would be of historical interest to bring my own personal testimony as an actor in that discovery.

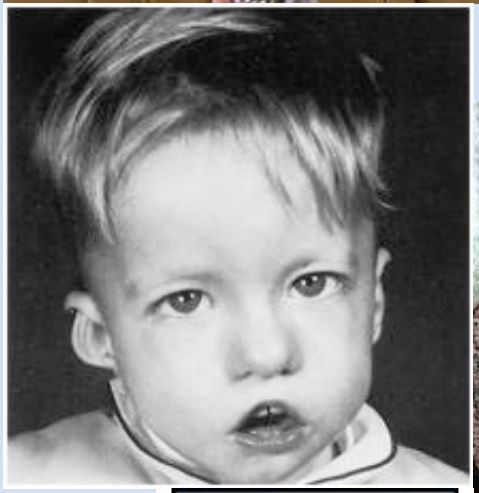
HISTORY OF SCIENCE

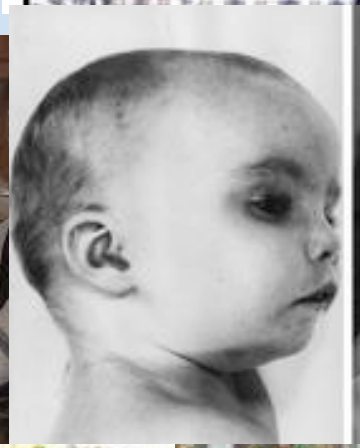
After More Than 50 Years, a Dispute Over Down Syndrome Discovery

Velocardiofacial (22q11.2) Syndrome











16p11.2 deletion, not in mother or father, only in child.

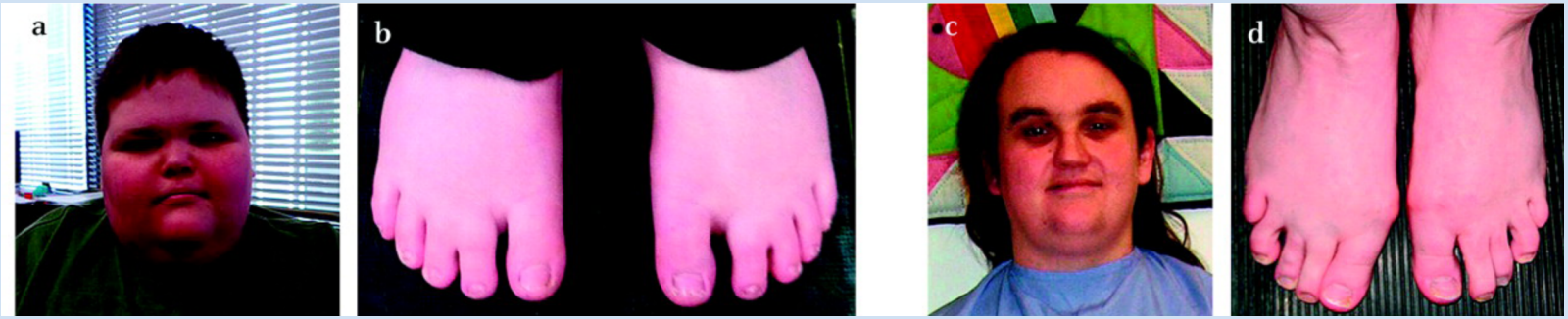
5 years old, but developmental age of 2 year old.
Speaks a few words, almost unintelligible.
Very hyperactive.
Can be withdrawn and has at times been diagnosed with "autism".

Current Diagnoses under Evaluation (DSM IV-TR)

AXIS I	299.00	Autism Disorder
	314.01	Attention-Deficit-Hyperactivity Disorder, Combined Type
AXIS II	V71.09	No Diagnosis
AXIS III	16p11.2	Microdeletion
AXIS IV		Psychosocial Stressors: Moderate(Adaptive/Behavioral and Educational/Learning Problems)
AXIS V		Current GAF: 60

*Private Photograph – Do not further distribute.

16p11.2 deletion



Clinical photographs. (a and b) Proband 2 (de novo deletion 16p11.2). Note long narrow palpebral fissures, short delicate nose, short neck and brachydactyly with 2–3 cutaneous toe syndactyly. (c and d) Mother of proband 3 (both with deletions). Note her large ears, smooth philtrum and short fifth toes.

Fernandez B A et al. J Med Genet 2010;47:195-203

JMG

Discovering a new syndrome and its genetic basis.

ARTICLE

Using VAAST to Identify an X-Linked Disorder Resulting in Lethality in Male Infants Due to N-Terminal Acetyltransferase Deficiency

Alan F. Rope,[1] Kai Wang,[2,19] Rune Evjenth,[3] Jinchuan Xing,[4] Jennifer J. Johnston,[5] Jeffrey J. Swensen,[6,7] W. Evan Johnson,[8] Barry Moore,[4] Chad D. Huff,[4] Lynne M. Bird,[9] John C. Carey,[1] John M. Opitz,[1,4,6,10,11] Cathy A. Stevens,[12] Tao Jiang,[13,14] Christa Schank,[8] Heidi Deborah Fain,[15] Reid Robison,[15] Brian Dalley,[16] Steven Chin,[6] Sarah T. South,[1,7] Theodore J. Pysher,[6] Lynn B. Jorde,[4] Hakon Hakonarson,[2] Johan R. Lillehaug,[3] Leslie G. Biesecker,[5] Mark Yandell,[4] Thomas Arnesen,[3,17] and Gholson J. Lyon[15,18,20,*]

The American Journal of Human Genetics *89*, 1–16, July 15, 2011

Published online 23 June 2011 | Nature | doi:10.1038/news.2011.382

News

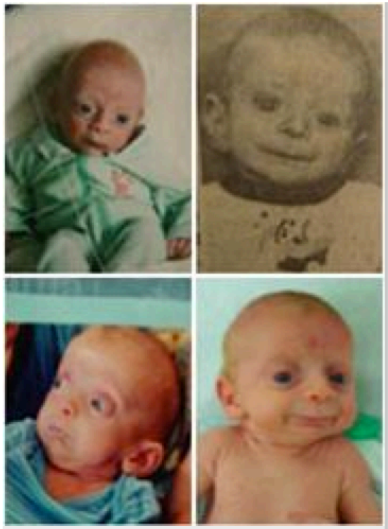
Software pinpoints cause of mystery genetic disorder

Genome analysis tools speedily track down previously unknown mutation.

Brendan Maher

Halena Black's first son, Kenny Rae, was born in November 1979. He struggled to put on weight, and had thin, wrinkled skin, big eyes and a broad mouth. In October the following year, he died from heart problems.

After Kenny Rae died, Black — a Mormon living in Ogden, Utah — had three healthy daughters before giving birth to another son in 1987. He had the same problem, and a similarly short lifespan. Her third son is healthy.



Four boys from the same family born wirh 'Ogden Syndrome'. Sufferers rarely survive for more than a year.

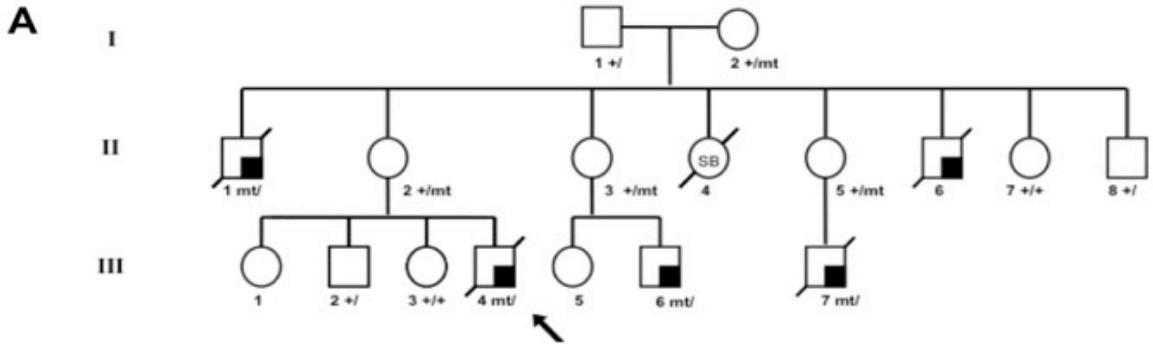
A. Rope et al./Am. J. Hum. Genet.

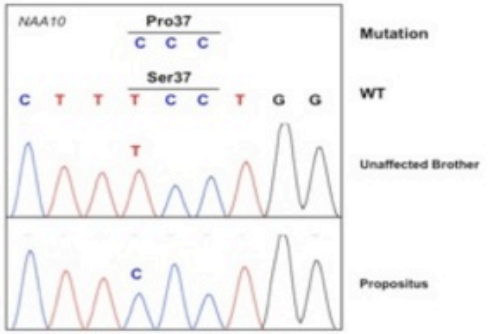
Black says she didn't dwell much on why her sons died until one of her daughters gave birth to a boy who looked just like Kenny Rae. "We didn't think that it passed on to the next generation. We didn't think that this would be a problem for them," says Black. All three daughters have since given birth to what the family calls 'little old men', one of whom died just last Sunday.

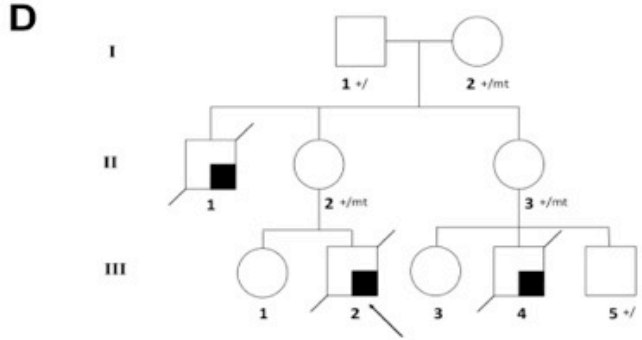
"This exemplifies an exceptionally rare disease, but the same type of strategy is now going to be applied to more common diseases to get the root cause," says Eric Topol, a medical geneticist at the Scripps Research Institute in La Jolla, California.

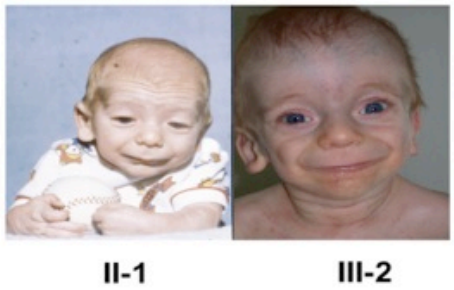
"This is one of the most exciting things in medicine," says Topol. "We're going to take the term 'idiopathic' which, basically means 'we don't' know,' and eliminate it."

Ogden Syndrome







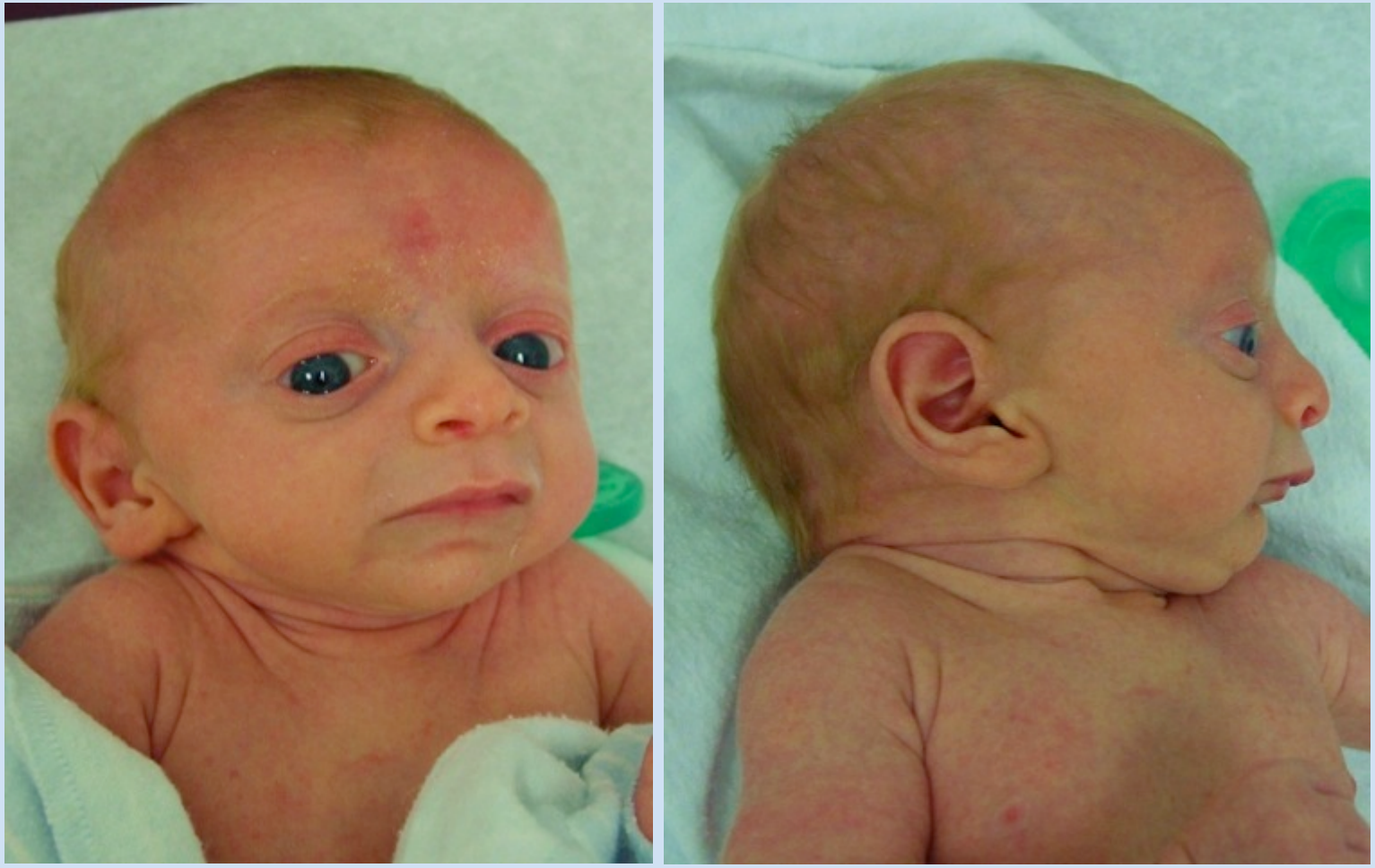


We found the SAME mutation in two unrelated families, with a very similar phenotype in both families, helping prove that this genotype contributes to the phenotype observed.

This is the first boy in the late 1970's.

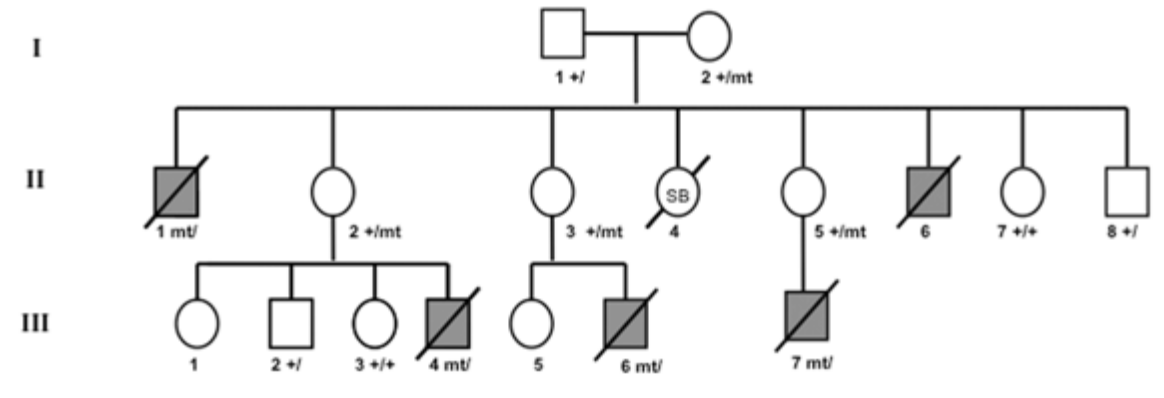
First boy. Called "a little old man" by the family. Died around ~1 year of age, from cardiac arrhythmias.

This is the "Proband" photograph presented at Case Conference.

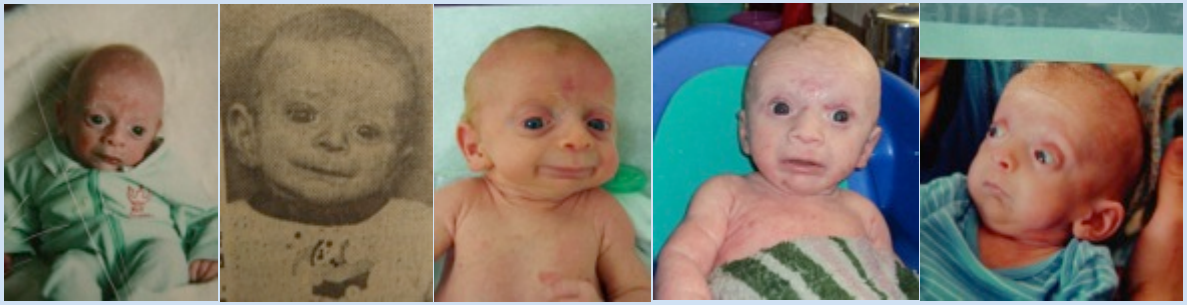


prominence of eyes, down-sloping palpebral fissures, thickened eyelids, large ears, beaking of nose, flared nares, hypoplastic nasal alae, short columella, protruding upper lip, micro-retrognathia

A

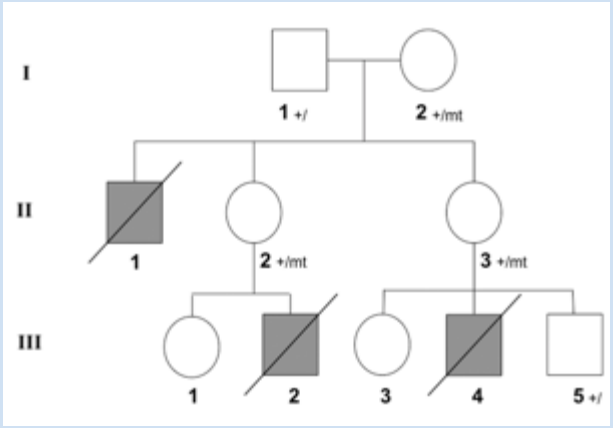


B



II-1 II-6 III-4 III-6 III-7

C



D

II-1 III-2

These are the Major Features of the Syndrome.

Table 1. Features of the syndrome	
Growth	post-natal growth failure
Development	global, severe delays
Facial	prominence of eyes, down-sloping palpebral fissures, thickened lids large ears beaking of nose, flared nares, hypoplastic alae, short columella protruding upper lip micro-retrognathia
Skeletal	delayed closure of fontanels broad great toes
Integument	redundancy / laxity of skin minimal subcutaneous fat cutaneous capillary malformations
Cardiac	structural anomalies (ventricular septal defect, atrial level defect, pulmonary artery stenoses) arrhythmias (Torsade de points, PVCs, PACs, SVtach, Vtach) death usually associated with cardiogenic shock preceded by arrythmia.
Genital	inguinal hernia hypo- or cryptorchidism
Neurologic	hypotonia progressing to hypertonia cerebral atrophy neurogenic scoliosis
Shaded regions include features of the syndrome demonstrating variability. Though variable findings of the cardiac, genital and neurologic systems were observed, all affected individuals manifested some pathologic finding of each.	

- **We performed X-chromosome exon capture with Agilent, followed by Next Gen Sequencing with Illumina.**

- **We analyzed the data with ANNOVAR and VAAST (Variant Annotation, Analysis and Search Tool). New computational tools for identifying disease-causing mutations by individual genome sequencing.**

Yandell, M. *et al.* 2011. "A probabilistic disease-gene finder for personal genomes." *Genome Res*. 21 (2011). doi:10.1101/gr.123158.111.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38, e164.

VAAST integrates AAS & Variant frequencies in a single probabilistic framework

- non-coding variants scored using allele frequency differences

- n_i : frequency of variant type among all variants observed in Background and Target genomes

- a_i: frequency of variant type among disease causing mutations in OMIM

- This approach means that *every* variant can be scored, non-synonymous, synonymous, coding, and non-coding. Phylogenetic conservation not required.

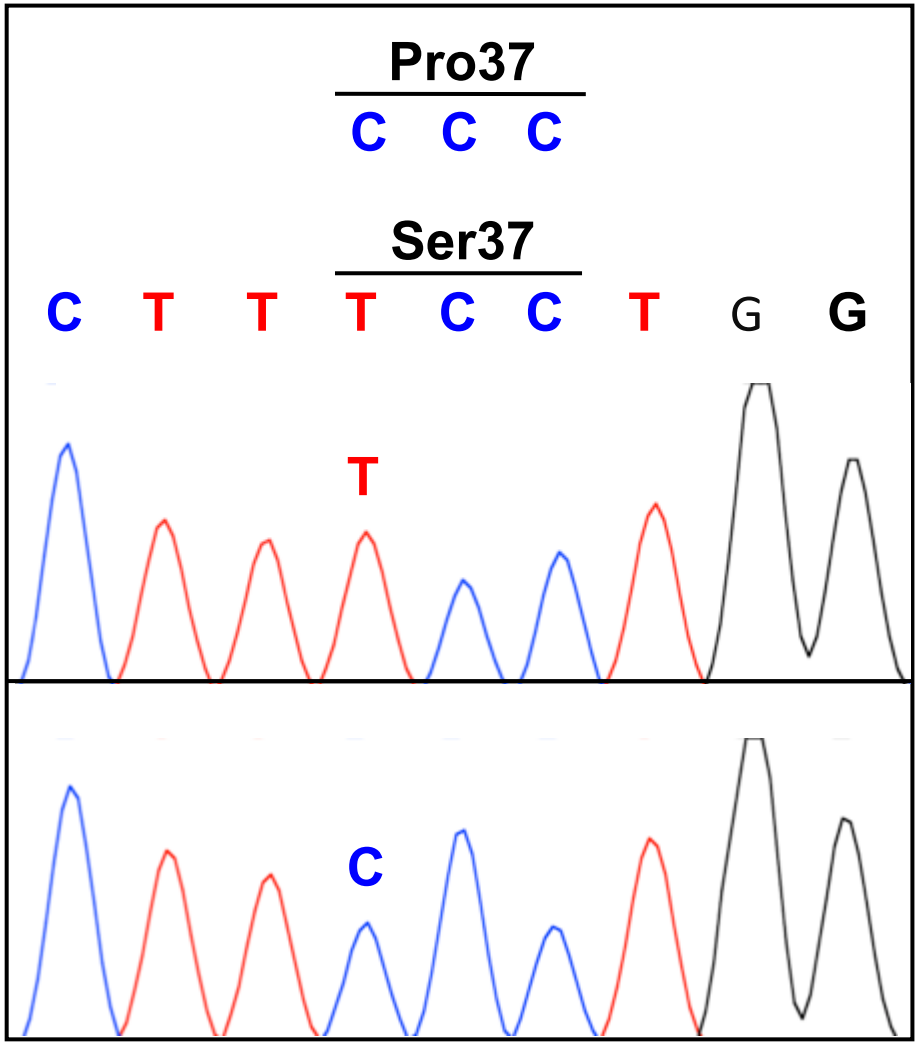
Omicia™ Unlocking Individualized Medicine

THE UNIVERSITY OF UTAH

Analysis with VAAST readily identified a few likely candidates.

Table 3. Summary of the filtering procedure and candidate genes using VAAST			
SNV calling pipeline	GATK	Samtools	GNUMAP
III-4 (total SNVs)	1546	1499	2168
III-4 (nsSNVs)	146	114	155
VAAST candidate genes (NAA10 ranking)	4 (3)	3 (2)	5 (2)
Present in III-4 and mother II-2 (nsSNVs)	122	107	116
VAAST candidate genes (NAA10 ranking)	3 (2)	2 (1)	2 (2)
Present in III-4, mother II-2, and grandmother I-2 (nsSNVs)	115	95	104
VAAST candidate genes (NAA10 ranking)	2 (1)	2 (1)	1 (1)
Present in III-4, II-2, and I-2, absent in brother III-2 and uncle II-8 (nsSNVs)	8	6	8
VAAST candidate genes (NAA10 ranking)	1 (1)	1 (1)	2 (1)

This is the mutation we found… one nucleotide change out of 6 billion nucleotides in a diploid genome…



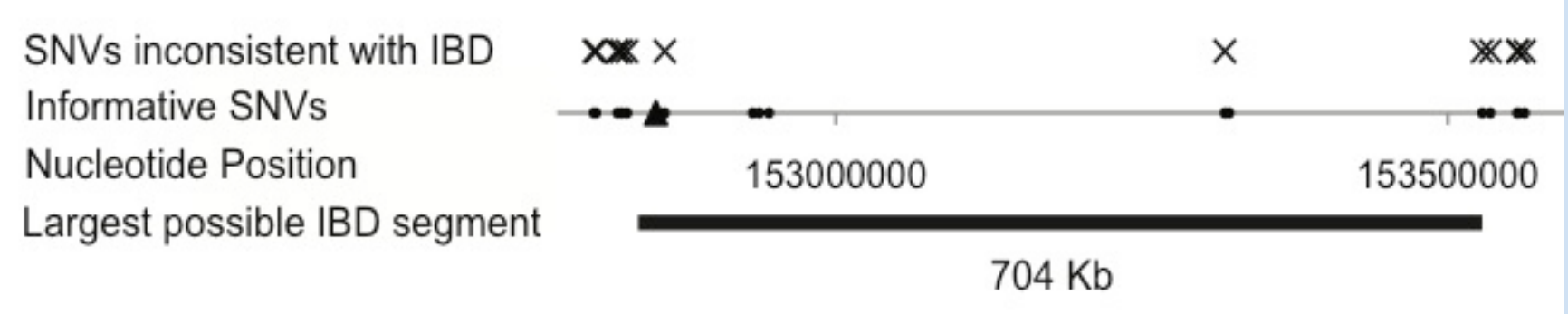
Mutation

WT

Unaffected Brother

Proband

Identity by Descent Analysis shows that the mutation must have arisen independently in two different families.



Courtesy of Chad Huff and Lynn Jorde

Prioritization of neurodevelopmental disease genes by discovery of new mutations

Alexander Hoischen[1], Niklas Krumm[2] & Evan E Eichler[2,3]

Advances in genome sequencing technologies have begun to revolutionize neurogenetics, allowing the full spectrum of genetic variation to be better understood in relation to disease. Exome sequencing of hundreds to thousands of samples from patients with autism spectrum disorder, intellectual disability, epilepsy and schizophrenia provides strong evidence of the importance of *de novo* and gene-disruptive events. There are now several hundred new candidate genes and targeted resequencing technologies that allow screening of dozens of genes in tens of thousands of individuals with high specificity and sensitivity. The decision of which genes to pursue depends on many factors, including recurrence, previous evidence of overlap with pathogenic copy number variants, the position of the mutation in the protein, the mutational burden among healthy individuals and membership of the candidate gene in disease-implicated protein networks. We discuss these emerging criteria for gene prioritization and the potential impact on the field of neuroscience.

Table 4 Recurrent identical *de novo* mutations in 6 genes identified in 11 exome studies with different neurodevelopmental phenotypes

Gene	Coding effect	Mutation (genomic DNA level)	Mutation (cDNA level)	Mutation (protein level)	Study	Disorder
ALG13	Missense	ChrX(GRCh37):g.110928268A>G	NM_001099922.2:c.320A>G	p.Asn107Ser	de Ligt *et al.*[1]	ID
ALG13	Missense	ChrX(GRCh37):g.110928268A>G	NM_001099922.2:c.320A>G	p.Asn107Ser	Allen *et al.*[11]	EE
ALG13	Missense	ChrX(GRCh37):g.110928268A>G	NM_001099922.2:c.320A>G	p.Asn107Ser	Allen *et al.*[11]	EE
KCNQ3	Missense	Chr8(GRCh37):g.133192493G>A	NM_001204824.1:c.328C>T	p.Arg110Cys	Rauch *et al.*[2]	ID
KCNQ3	Missense	Chr8(GRCh37):g.133192493G>A	NM_001204824.1:c.328C>T	p.Arg110Cys	Allen *et al.*[11]	EE
SCN1A	Splice donor	LRG_8:g.24003G>A	NM_006920.4:c.602+1G>A	p.?	Allen *et al.*[11]	EE
SCN1A	Splice donor	LRG_8:g.24003G>A	NM_006920.4:c.602+1G>A	p.?	Allen *et al.*[11]	EE
CUX2	Missense	Chr12(GRCh37):g.111748354G>A	NM_015267.3:c.1768G>A	p.Glu590Lys	Rauch *et al.*[2]	ID
CUX2	Missense	Chr12(GRCh37):g.111748354G>A	NM_015267.3:c.1768G>A	p.Glu590Lys	Allen *et al.*[11]	EE
SCN2A	Missense	Chr2(GRCh37):g.166198975G>A	NM_021007.2:c.2558G>A	p.Arg853Gln	Allen *et al.*[11]	EE
SCN2A	Missense	Chr2(GRCh37):g.166198975G>A	NM_021007.2:c.2558G>A	p.Arg853Gln	Allen *et al.*[11]	EE
DUSP15	Missense	Chr20(GRCh37):g.30450489G>A	NM_080611.2:c.320C>T	p.Thr107Met	Neale *et al.*[7]	ASD
DUSP15	Missense	Chr20(GRCh37):g.30450489G>A	NM_080611.2:c.320C>T	p.Thr107Met	Fromer *et al.*[10]	SCZ

EE, epileptic encephalopathies; ASD, autism spectrum disorder; ID, intellectual disability; SCZ, schizophrenia.

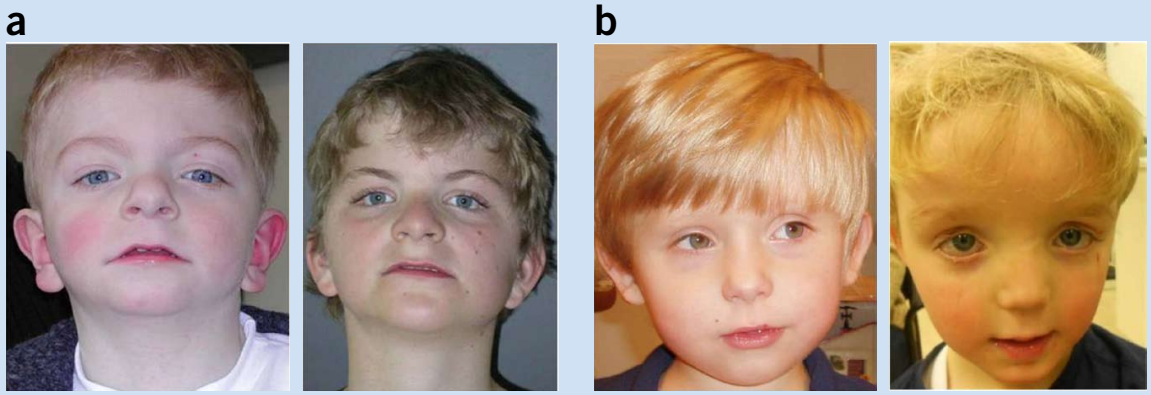
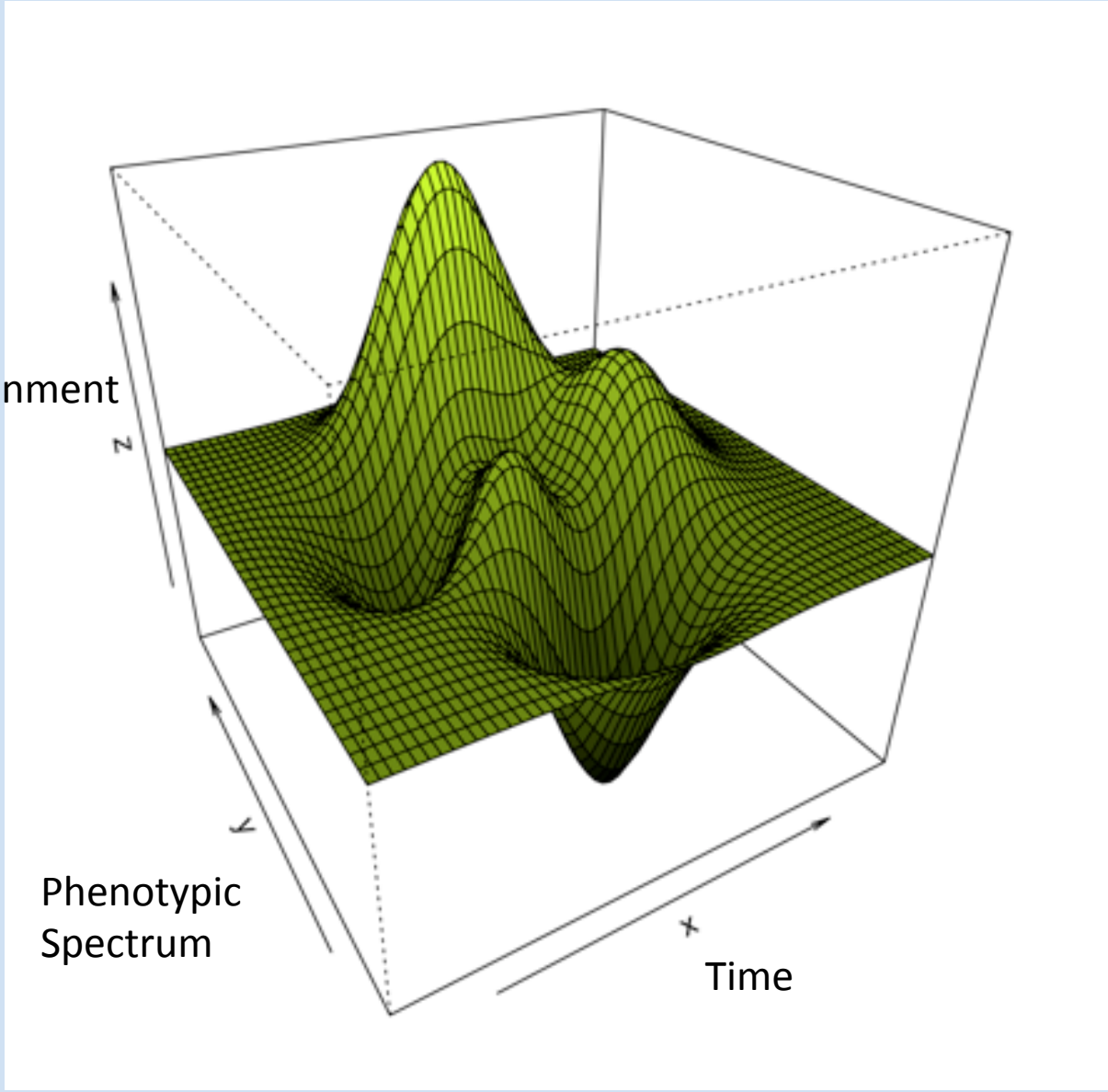


Figure 3 Phenotypic similarity of two patients with identical *PACS1 de novo* mutations and two patients with similar *ADNP* mutations. (**a**) These two unrelated patients show identical *de novo* point mutations (c.607C>T; p.Arg203Trp) in *PACS1* (RefSeq NM_018026.3)[53]. The striking similarity in phenotype includes low anterior hairline, highly arched eyebrows, synophrys, hypertelorism with downslanted palpebral fissures, long eyelashes, a bulbous nasal tip, a flat philtrum with a thin upper lip, downturned corners of the mouth and low-set ears. Reprinted from ref. 53, Copyright (2012), with permission from The American Society of Human Genetics. (**b**) These two unrelated patients both show LoF mutations in *ADNP* (c.2496_2499delTAAA; p.Asp832Lysfs*80 and c.2157C>G; p.Tyr719*)[44] resulting in a new SWI-SNF–related autism syndrome. Patients present with clinical similarities, including a prominent forehead, a thin upper lip and a broad nasal bridge. Reprinted from ref. 44.

Clinical genetics of neurodevelopmental disorders

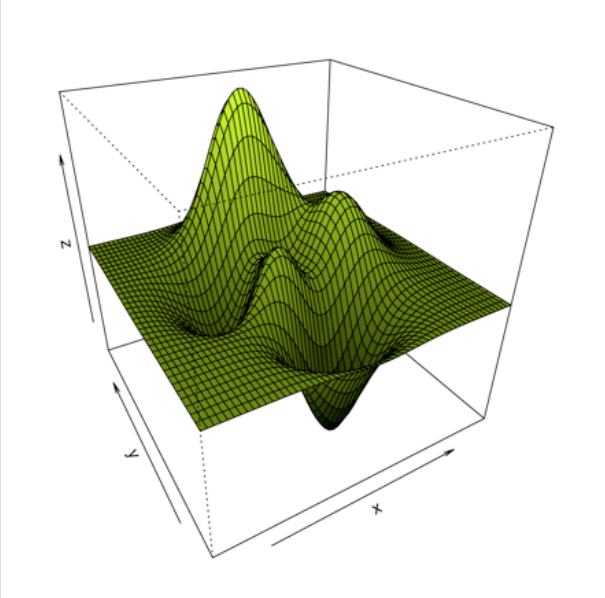
Gholson J Lyon and Jason O'Rawe

bioRxiv posted online November 18, 2013
Access the most recent version at doi:10.1101/000687



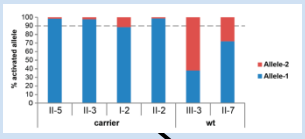
A conceptual model of genotype-phenotype correlations. The *y* plane represents a phenotypic spectrum, the *x* plane represents the canalized progression of development through time, and the *z* plane represents environmental fluctuations.

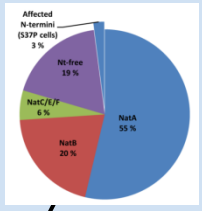
Once one finds a validated high-effect size mutation, functional analysis is appropriate.

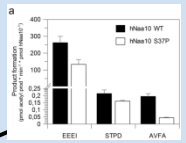


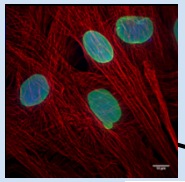
patient cells

molecular level









X-chromosome skewing in carriers

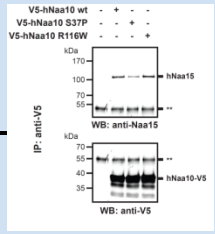
affected *in vivo* substrates

reduced enzymatic activity

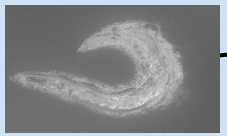
normal nuclear envelope

Naa10 S37P

disrupted NatA complex



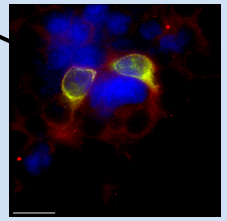
C. elegans



initial experiments

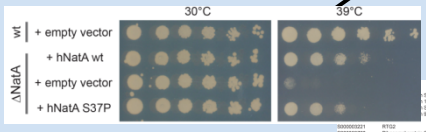
HEK293 cells

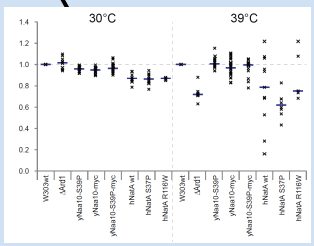
normal localization



growth defect at 39°C
ribosome defect?
no difference cyclo/doxo?

S39 mutation?
growth defect of "humanized" strains





S. cerevisiae

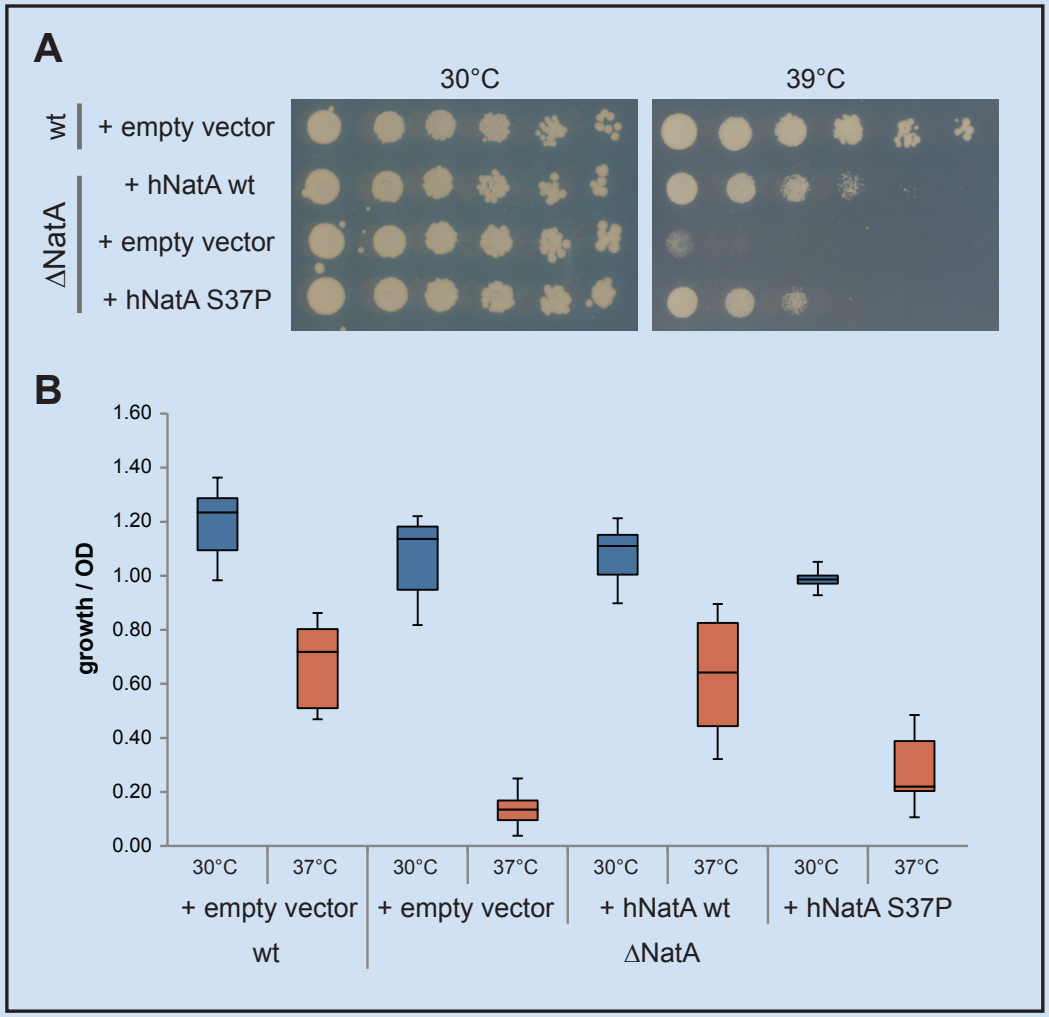
Gene	Length		Gene	Length	Identity Score %
human	235	vs	c_elegans	181	68
human	235	vs	s_pombe	177	57
human	235	vs	s_cerevisiae	238	38
c_elegans	181	vs	s_pombe	177	57
c_elegans	181	vs	s_cerevisiae	238	48
s_pombe	177	vs	s_cerevisiae	238	52

The Scores Table shows the pairwise scores calculated for every pair of sequences that is to be aligned. Pairwise scores are simply the number of identities between the two sequences, divided by the length of the alignment, and represented as a percentage. This alignment is only a precursor to the full multiple alignment and might not be preserved.

```
CLUSTAL 2.1 multiple sequence alignment

h_sapiens_V1      --MNIRNARPEDLMNMQHCNLLCLPENYQMKYYFYHGLSWPQLSYIAE------------ 46
c_elegans         --MNIRCARVDDLMSMQNANLMCLPENYQMKYYFYHALSWPQLSYIAE------------ 46
s_pombe           --MDIRPARISDLTGMQNCNLHNLPENYQLKYYLYHAISWPMLSYVAT------------ 46
s_cerevisiae      MPINIRRATINDIICMQNANLHNLPENYMMKYYMYHILSWPEASFVATTTTLDCEDSDEQ 60
                    ::** *  .*:  **:.**  ***** :***:** :***  *::*

h_sapiens_V1      DENG--------------------------KIVGYVLAKMEEDPDD---VPHGHITSLAV 77
c_elegans         DHKG--------------------------NVVGYVLAKMEEDPGE---EPHGHITSLAV 77
s_pombe           DPKG--------------------------RVVGYVLAKMEEEPKDG--IPHGHITSVSV 78
s_cerevisiae      DENDKLELTLDGTNDGRTIKLDPTYLAPGEKLVGYVLVKMNDDPDQQNEPPNGHITSLSV 120
                  * :.                          .:*****.**:::* :    *:******::*

h_sapiens_V1      KRSHRRLGLAQKLMDQASRAMIENFNAKYVSLHVRKSNRAALHLYSNTLNFQISEVEPKY 137
c_elegans         KRSYRRLGLANKMMDQTARAMVETYNAKYVSLHVRVSNRAALN-YKNTLKFEIVDTEPKY 136
s_pombe           MRSYRHLGLAKRLMVQSQRAMVEVYGAKYMSLHVRKSNRAAIHLYRDTLQFDVQGIESKY 138
s_cerevisiae      MRTYRRMGIAENLMRQALFALREVHQAEYVSLHVRQSNRAALHLYRDTLAFEVLSIEKSY 180
                   *::*::* :*:.:* *:  *: * . *:*:***** *****:: * :** *::   * .*

h_sapiens_V1      YADGEDAYAMKR--DLTQMADELRRHLELKEKGRHVVLGAIENKVESKGNSPPSSGEACR 195
c_elegans         YADGEDAYAMRR--DLAKWAEE--RNIEPADR---------------------------- 164
s_pombe           YADGEDAYAMHK--DFSTLKFD---TPETN------------------------------ 163
s_cerevisiae      YQDGEDAYAMKKVLKLEELQISNFTHRRLKENE---------------------------- 213
                  * ********::  .:     .     .

h_sapiens_V1      EEKGLAAEDSGGDSKDLSEVSETTESTDVKDSSEASDSAS-- 235
c_elegans         ------------------EAYTTAKTTDDKKKNRS------- 181
s_pombe           -----------------DELAKTVQSLALNN----------- 177
s_cerevisiae      -----------------EKLEDDLESDLLEDIIKQGVNDIIV 238
                                  :     ::   :.
```

yeast growth
YPDA media (Clonetech, #630464)
Yeast minimal SD base (Clonetech, #630411) supplemented with drop out mix –Ura (Clonetech, #630416)

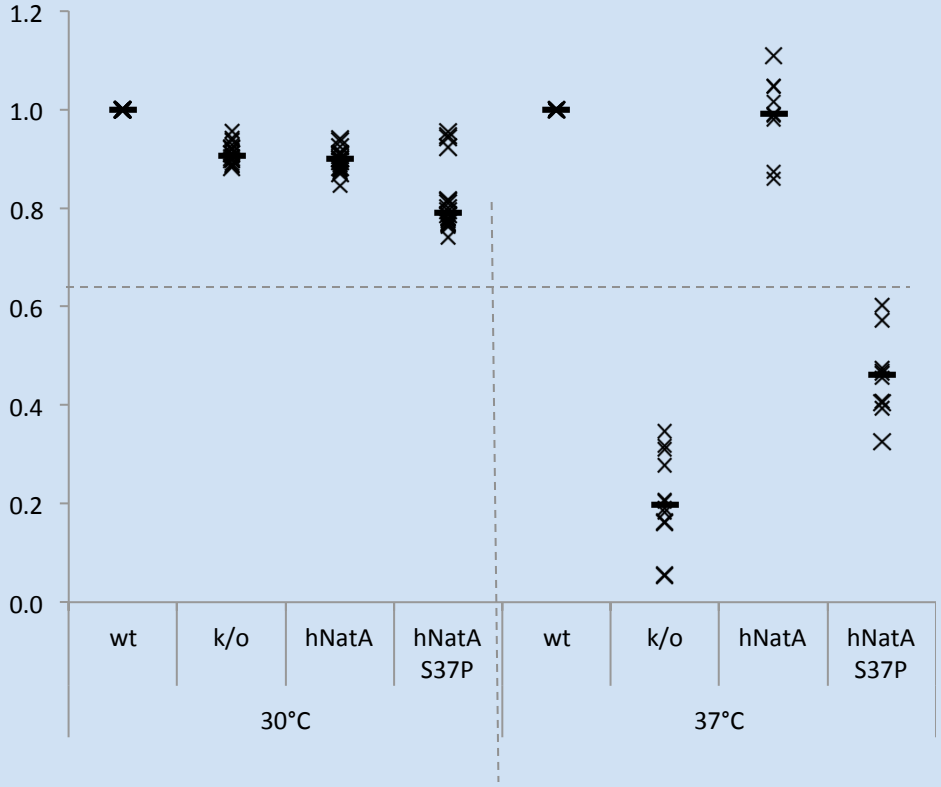
A 5 ml overnight culture was grown in SD^{-URA} at 30°C. Cells were diluted to an OD_{600} of 0.1 and either spotted in 1:5 serial dilutions on plates for 48 h (upper panel) or grown in 2 ml cultures at 30°C or 39°C under constant agitation for 24 h (lower panel). Optical density was plotted
n=11



Same data, just presented differently.

yeast growth
YPDA media (Clonetech, #630464)
Yeast minimal SD base (Clonetech, #630411) supplemented with drop out mix –Ura (Clonetech, #630416)

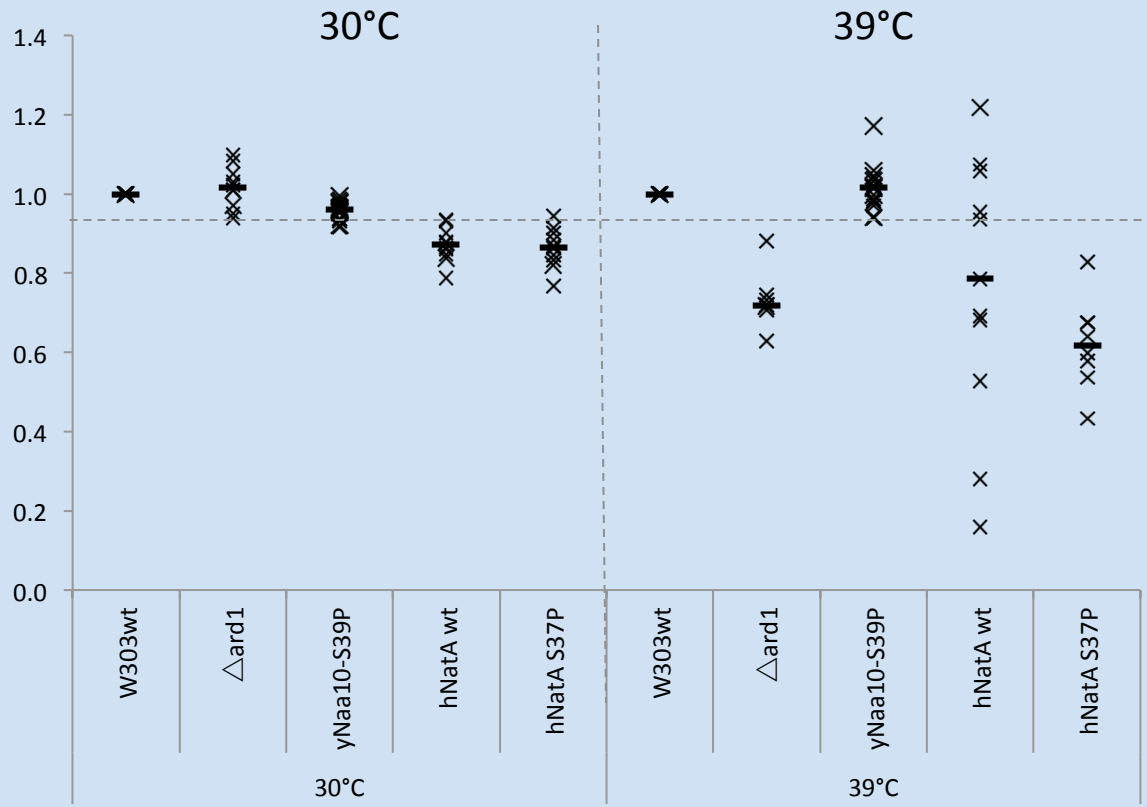
A 5 ml overnight culture was grown in SD^{-URA} at 30°C. Cells were diluted to an OD_{600} of 0.1 and either spotted in 1:5 serial dilutions on plates for 48 h (upper panel) or grown in 2 ml cultures at 30°C or 39°C under constant agitation for 24 h (lower panel). Optical density was plotted
n=11



Endogenous, single-copy genes in yeast.

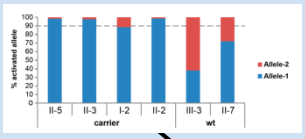
Optical density as a measure of growth was normalized to the W303 wt strain for every independent experiment and plotted (X). The median of all experiments is shown as a short line

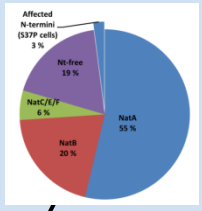
n=22 for S39P
n=11 for all other strains

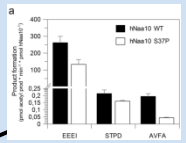


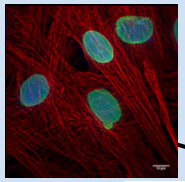
patient cells

molecular level









X-chromosome skewing in carriers

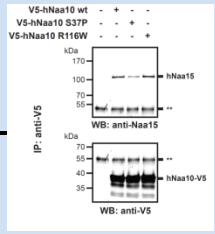
affected *in vivo* substrates

reduced enzymatic activity

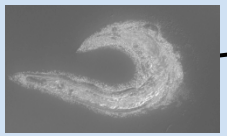
normal nuclear envelope

Naa10 S37P

disrupted NatA complex



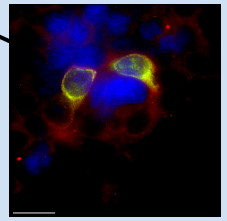
C. elegans



initial experiments

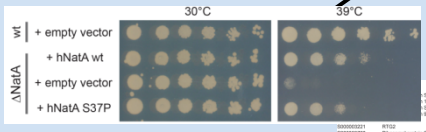
HEK293 cells

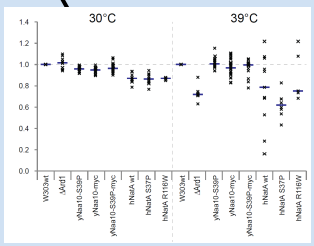
normal localization



growth defect at 39°C
ribosome defect?
no difference cyclo/doxo?

S39 mutation?
growth defect of "humanized" strains





S. cerevisiae

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Contemporary, yeast-based approaches to understanding human genetic variation

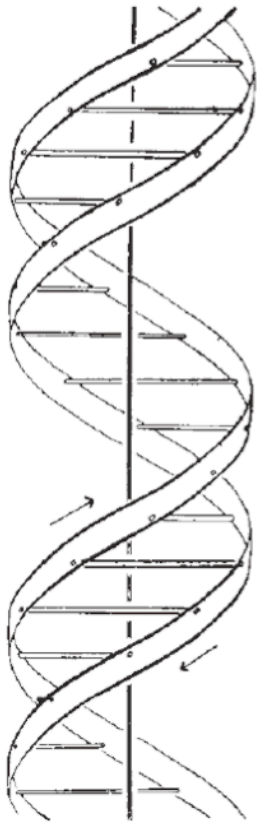
Maitreya J Dunham and Douglas M Fowler

Learning Objectives

- The participants will be able to discuss the extraordinary amount of variable expressivity seen in neurodevelopmental disorders.

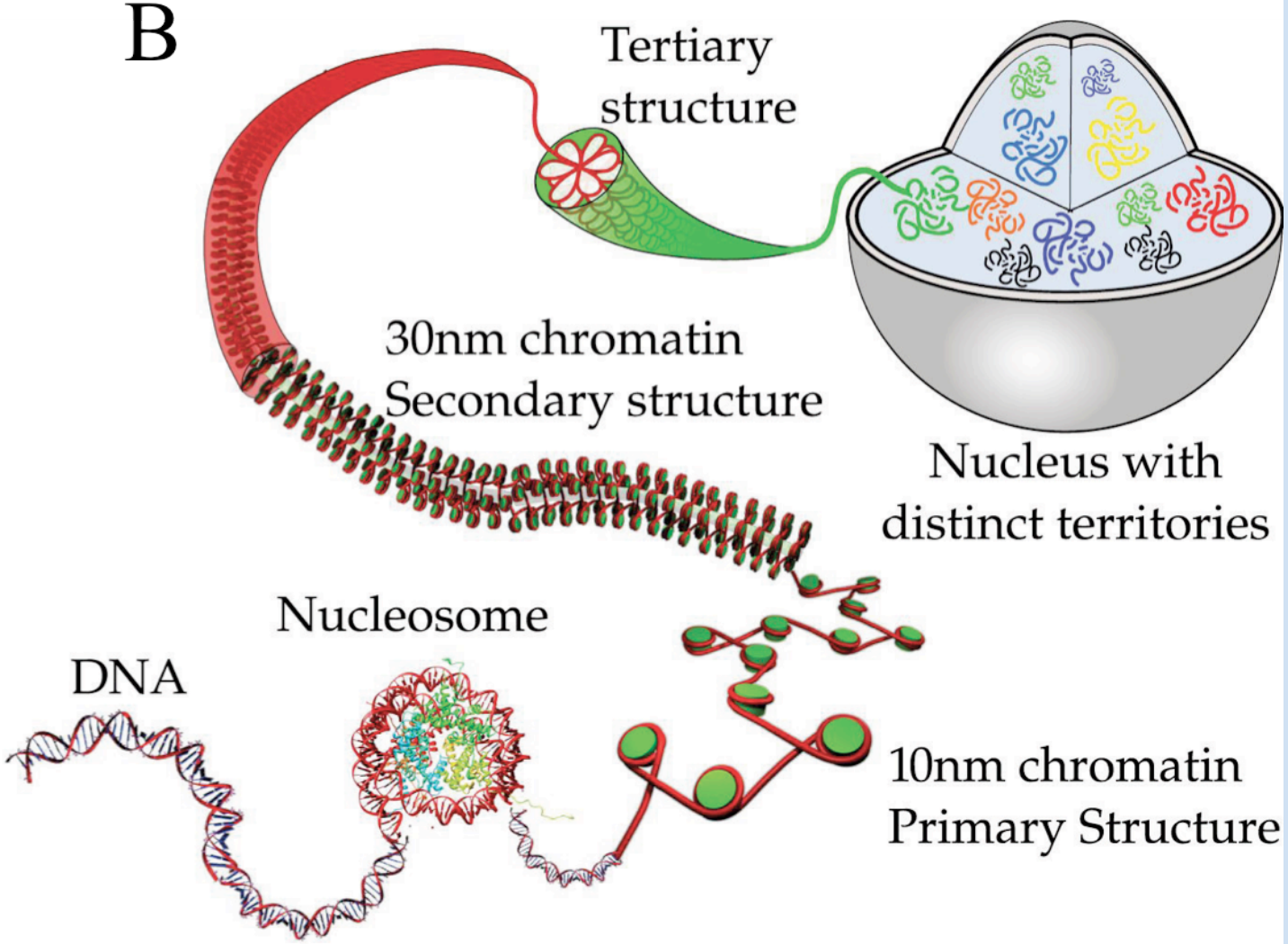
- **The participants will plan ways in which to integrate genomic and phenotypic longitudinal data to prevent the development of certain illnesses.**

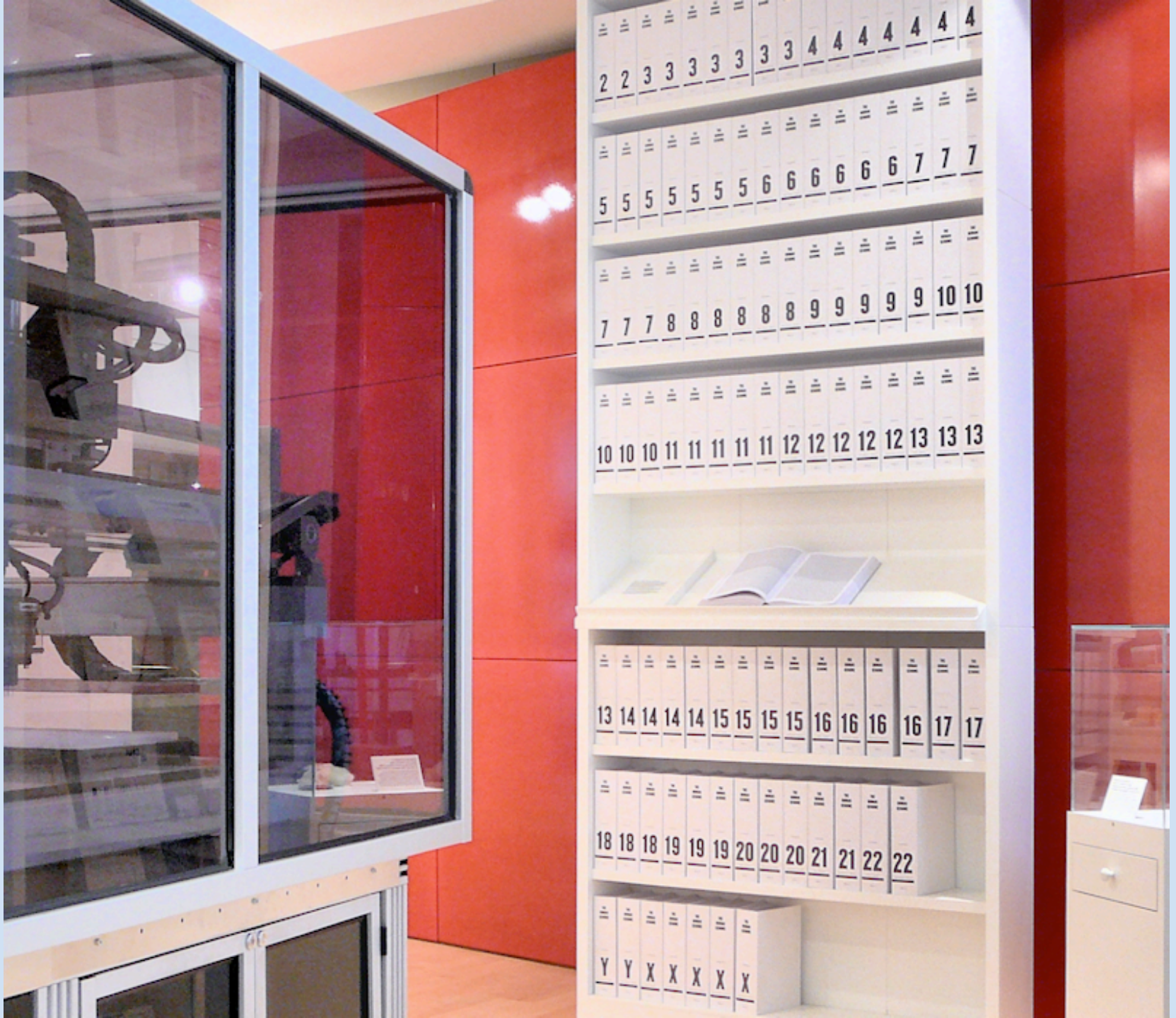
A

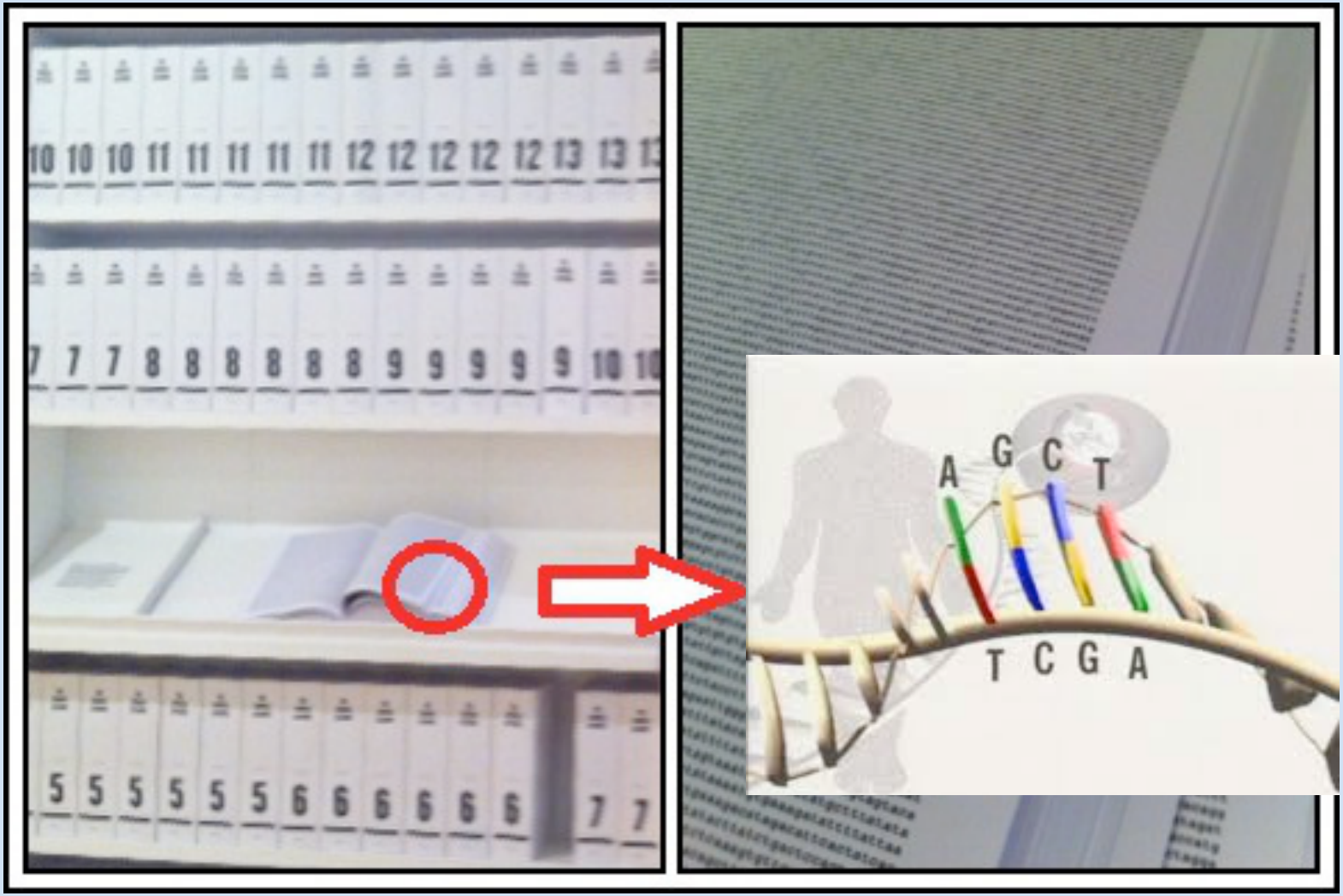


This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

B

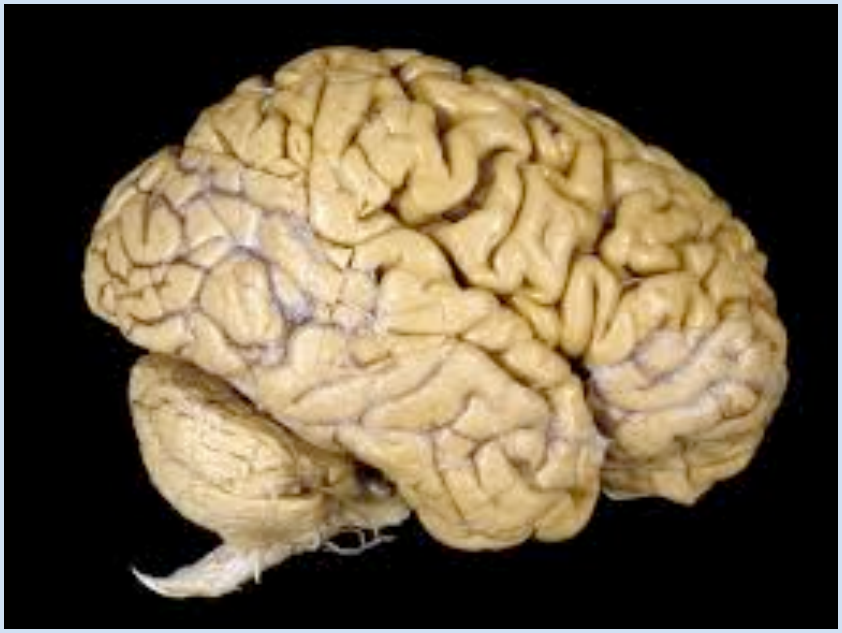





A G C T
T C G A

In the year 2014….
This:

Is orders of magnitudes easier than this:



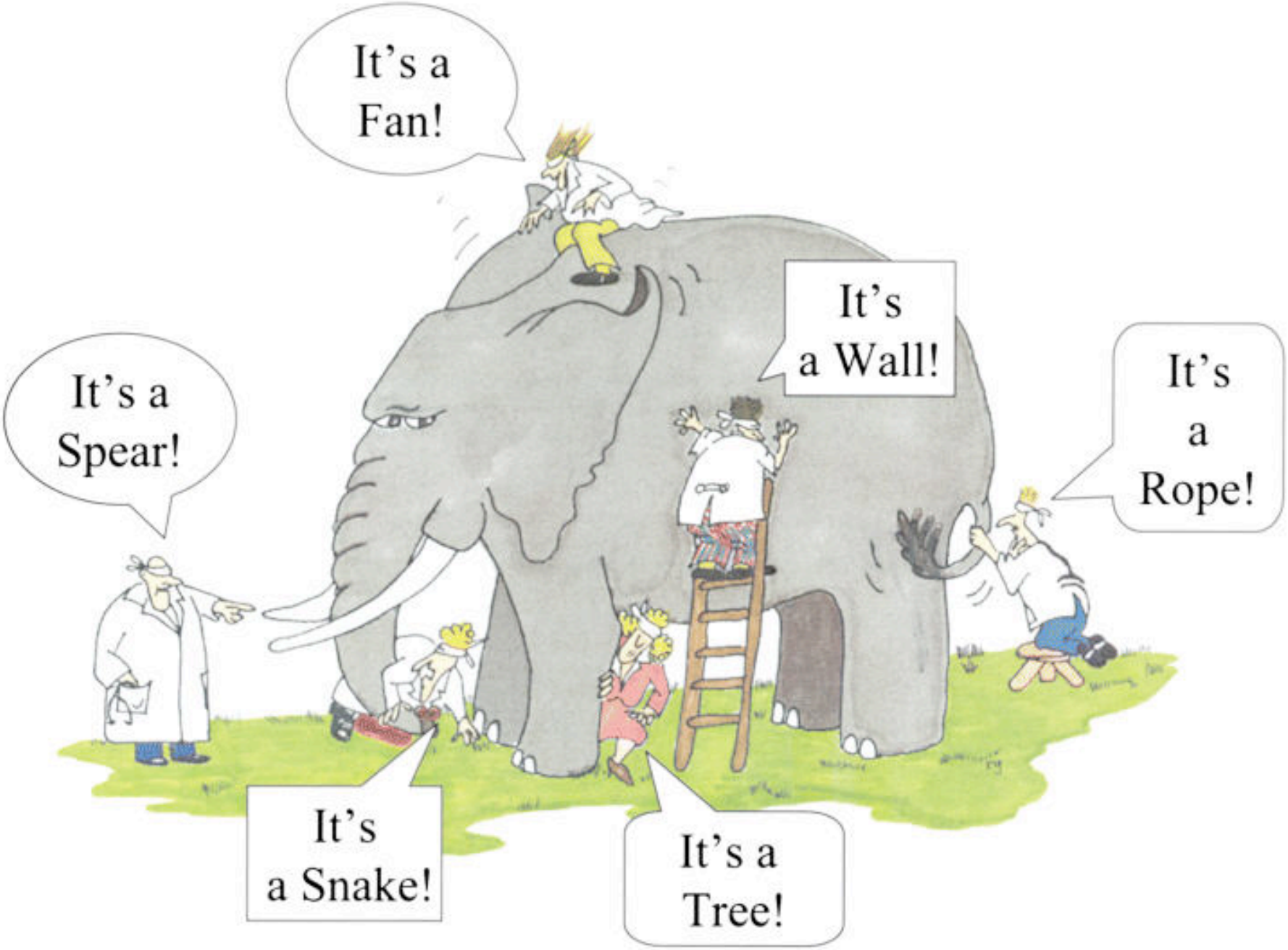
BRAIN 2025

A SCIENTIFIC VISION

Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Working Group Report to the Advisory Committee to the Director, NIH

June 5, 2014

NIH National Institutes of Health
Turning Discovery into Health


It's a Fan!
It's a Spear!
It's a Wall!
It's a Rope!
It's a Snake!
It's a Tree!

And yet, remarkably:

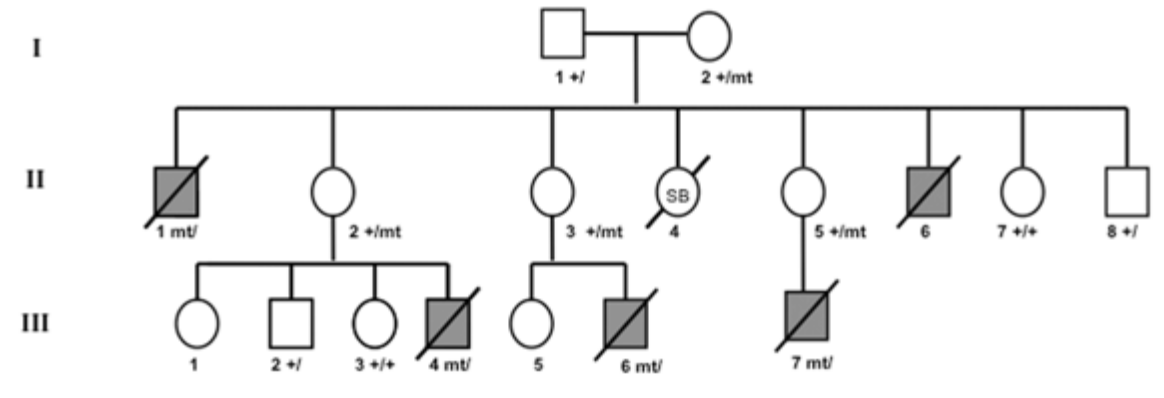
- We do not really know the expression of pretty much ALL mutations in **humans**, as we have not systematically sequenced or karyotyped any genetic alteration in **Thousands to Millions** of **randomly** selected people.

A Genotype-First Approach to Defining the Subtypes of a Complex Disease

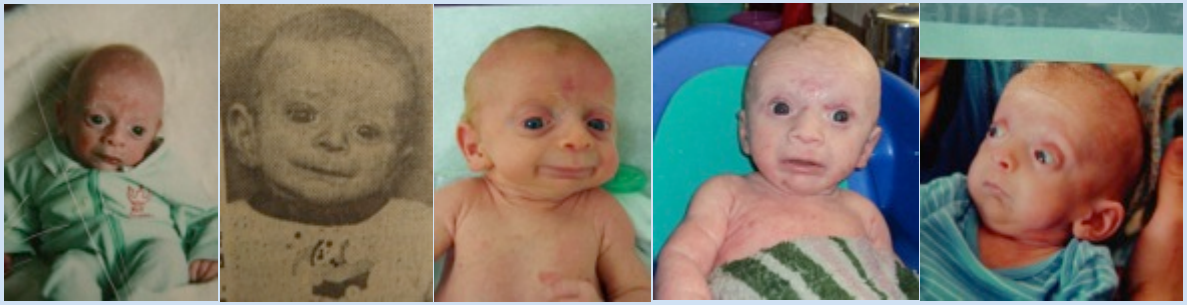
Holly A. Stessman,[1] Raphael Bernier,[2] and Evan E. Eichler[1,3,*]
[1]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
[2]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA 98195, USA
[3]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA
*Correspondence: eee@gs.washington.edu
http://dx.doi.org/10.1016/j.cell.2014.02.002

Medical genetics typically entails the detailed characterization of a patient's phenotypes followed by genotyping to discover the responsible gene or mutation. Here, we propose that the systematic discovery of genetic variants associated with complex diseases such as autism are progressing to a point where a reverse strategy may be fruitful in assigning the pathogenic effects of many different genes and in determining whether particular genotypes manifest as clinically recognizable phenotypes. This "genotype-first" approach for complex disease necessitates the development of large, highly integrated networks of researchers, clinicians, and patient families, with the promise of improved therapies for subsets of patients.

A

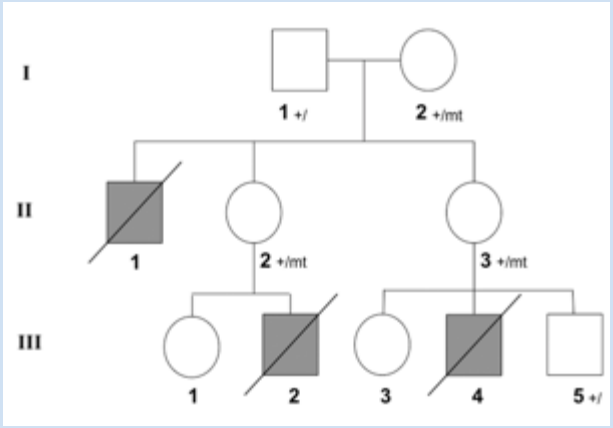


B



II-1 II-6 III-4 III-6 III-7

C



D

II-1 III-2

Published Ahead of Print on April 14, 2014 as 10.1200/JCO.2013.53.6607
The latest version is at http://jco.ascopubs.org/cgi/doi/10.1200/JCO.2013.53.6607

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Clinical Evaluation of a Multiple-Gene Sequencing Panel for Hereditary Cancer Risk Assessment

Allison W. Kurian, Emily E. Hare, Meredith A. Mills, Kerry E. Kingham, Lisa McPherson, Alice S. Whittemore, Valerie McGuire, Uri Ladabaum, Yuya Kobayashi, Stephen E. Lincoln, Michele Cargill, and James M. Ford

Processed as a Rapid Communication manuscript

INVITAE
Reinventing genetic testing.

Major barriers to the widespread implementation of genomic medicine in the clinic.

- Limits of our current technology & knowledge
- Lack of public education
- Lack of physician knowledge about genetics
- Apathy on the part of the populace in terms of preventive efforts
- Reluctance of insurance companies & governments to pay for genetic testing
- Focus in our society on treatment, not on early diagnosis and prevention
- Privacy concerns

Lyon and Wang *Genome Medicine* 2012, **4**:58
http://genomemedicine.com/content/4/7/58

Genome Medicine

REVIEW

Identifying disease mutations in genomic medicine settings: current challenges and how to accelerate progress

Gholson J Lyon[*1,2] and Kai Wang[*2,3]

"It is perhaps naive to expect that these obstacles can be overcome within the next 20 years, and it may very well be the case that there might be a 50-year time horizon on the secure implementation of clinical genomics and individualized medicine. We certainly hope that every newborn will have the vast majority of their genome sequenced and digitally available by the year 2062".

Limits of our current technology & knowledge

Analytic Validity

- Sequencing “clinical-grade genomes”
- Bioinformatics analysis

Clinical Validity

- Genetic architecture of illness

Limits of our current technology & knowledge

Analytic Validity

- **Sequencing "clinical-grade genomes"**

- Bioinformatics analysis

Clinical Validity

- Genetic architecture of illness

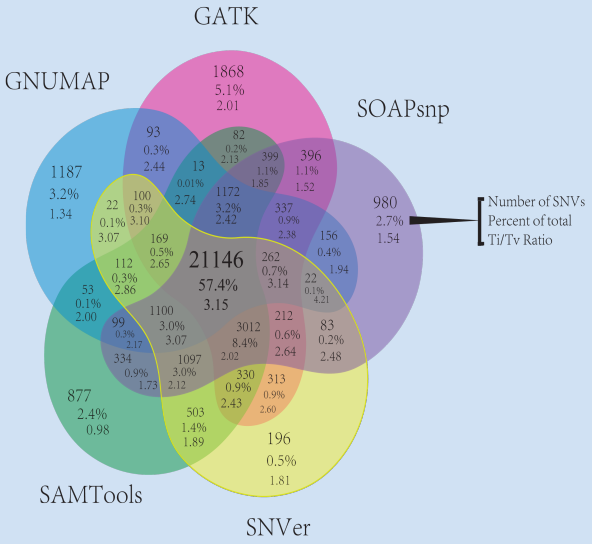
Genome Medicine

RESEARCH Open Access

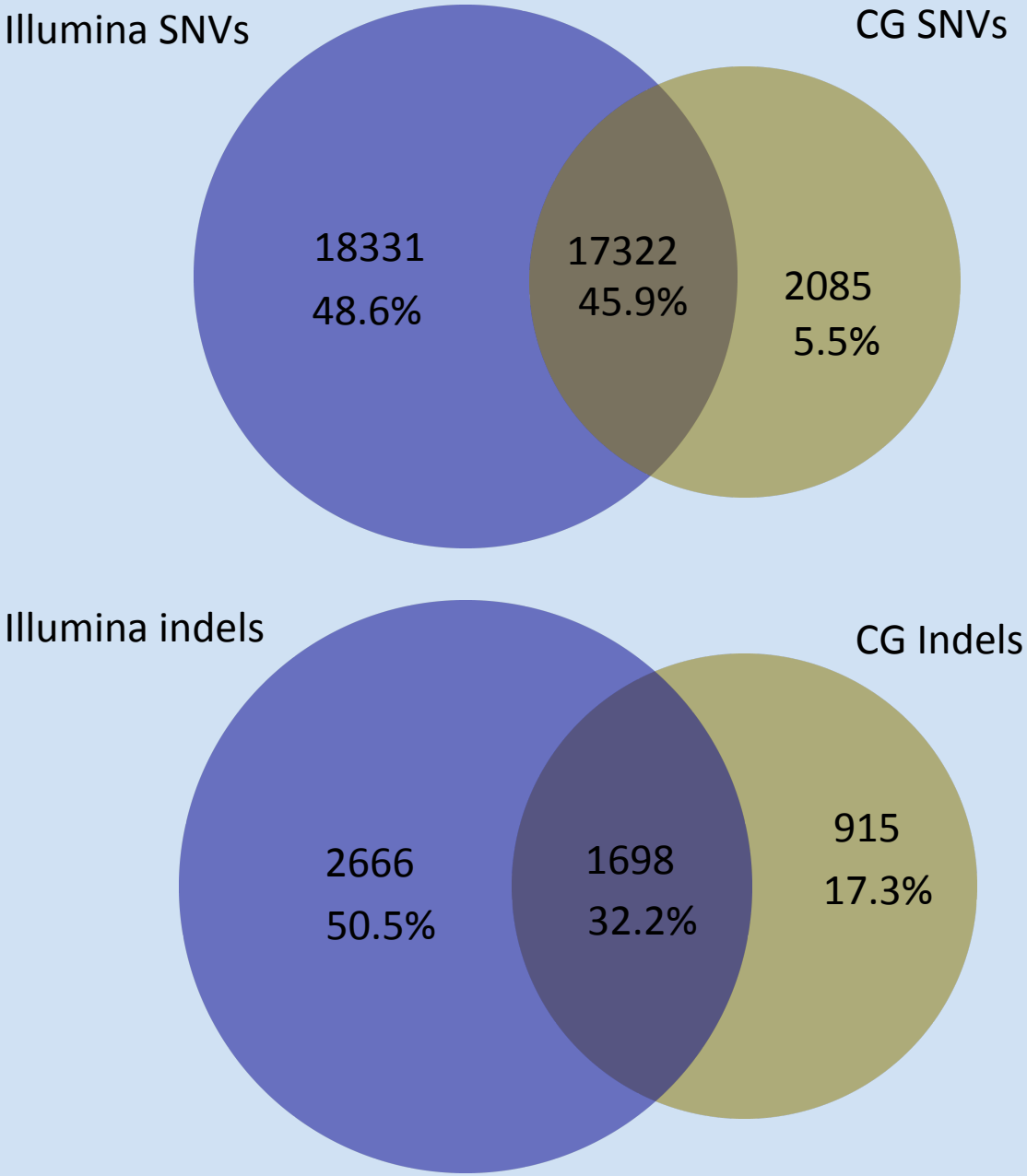
Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing

Jason O'Rawe[1,2], Tao Jiang[3], Guangqing Sun[3], Yiyang Wu[1,2], Wei Wang[4], Jingchu Hu[3], Paul Bodily[5], Lifeng Tian[6], Hakon Hakonarson[6], W Evan Johnson[7], Zhi Wei[4], Kai Wang[8,9*] and Gholson J Lyon[1,2,9*]

Conclusions: Our results suggest that more caution should be exercised in genomic medicine settings when analyzing individual genomes, including interpreting positive and negative findings with scrutiny, especially for indels. We advocate for renewed collection and sequencing of multi-generational families to increase the overall accuracy of whole genomes.



What is the “True” Personal Genome?



Limits of our current technology & knowledge

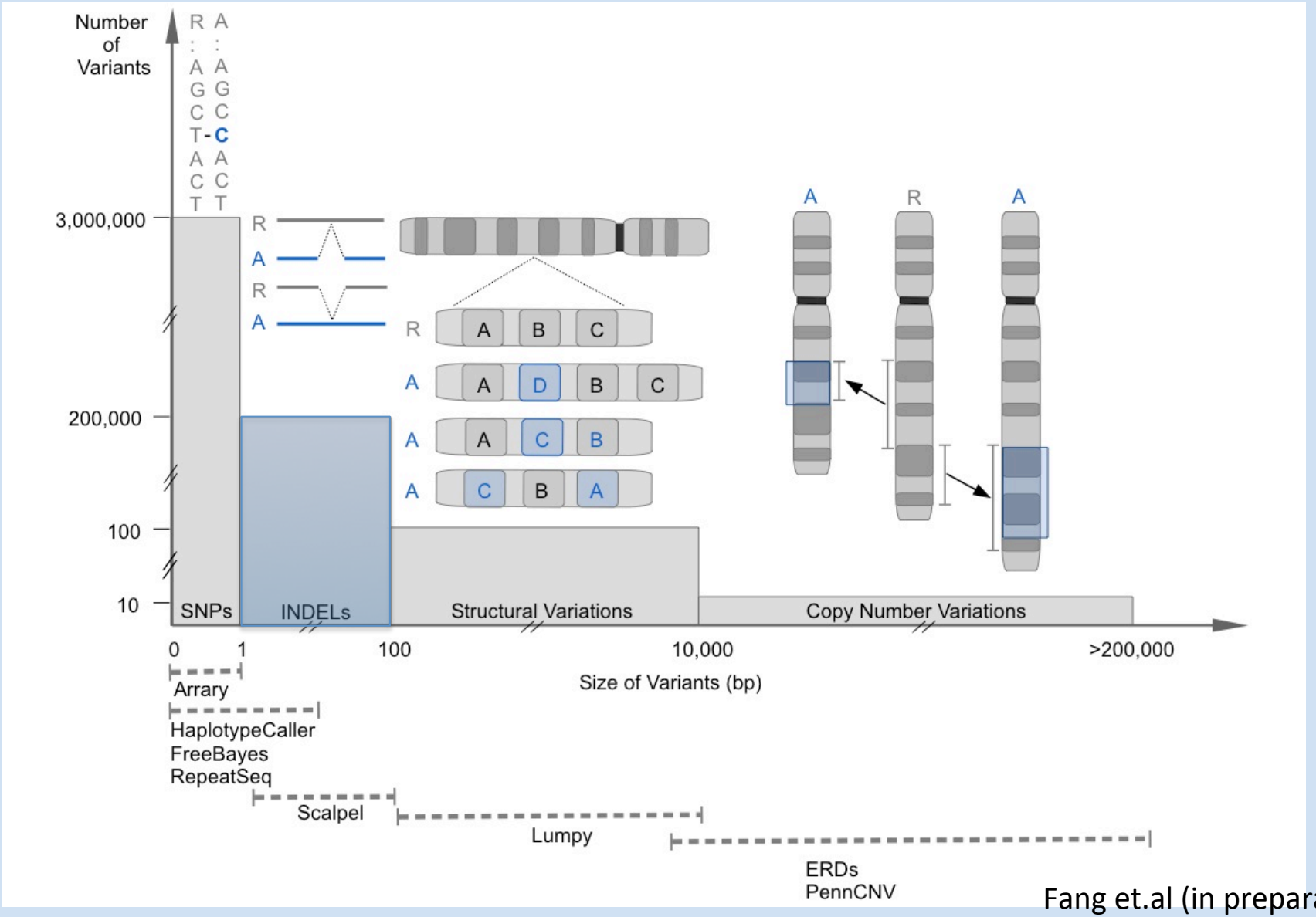
Analytic Validity

- Sequencing “clinical-grade genomes”
- **Bioinformatics analysis**

Clinical Validity

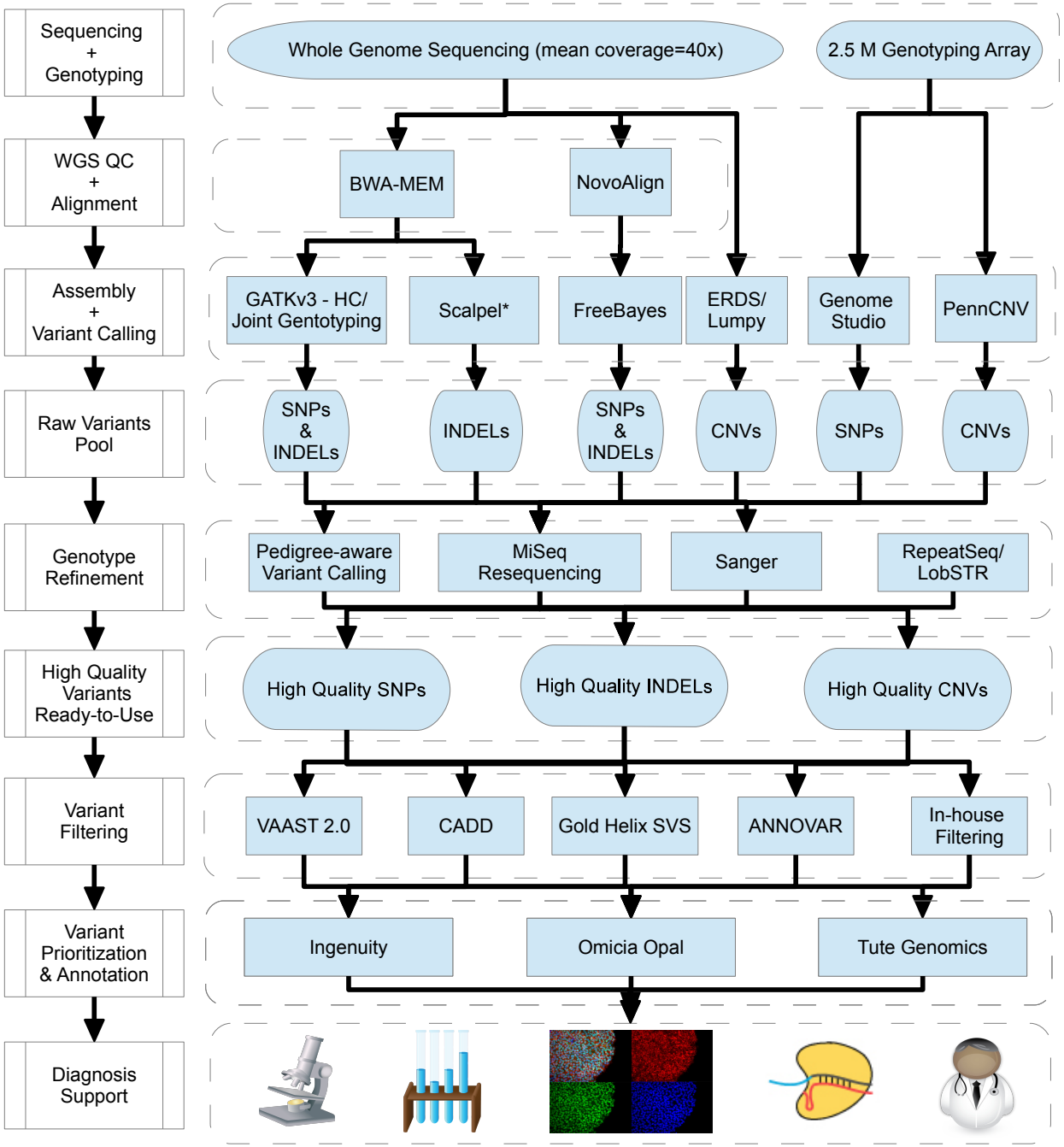
- Genetic architecture of illness

Interrogating human genome from single-codon resolution to large structural events with WGS



Fang et.al (in preparation)

Variant Analysis Pipeline for Whole Genome Sequencing Data



* Scalpel (In press) http://schatzlab.cshl.edu/

bioRχiv beta
THE PREPRINT SERVER FOR BIOLOGY

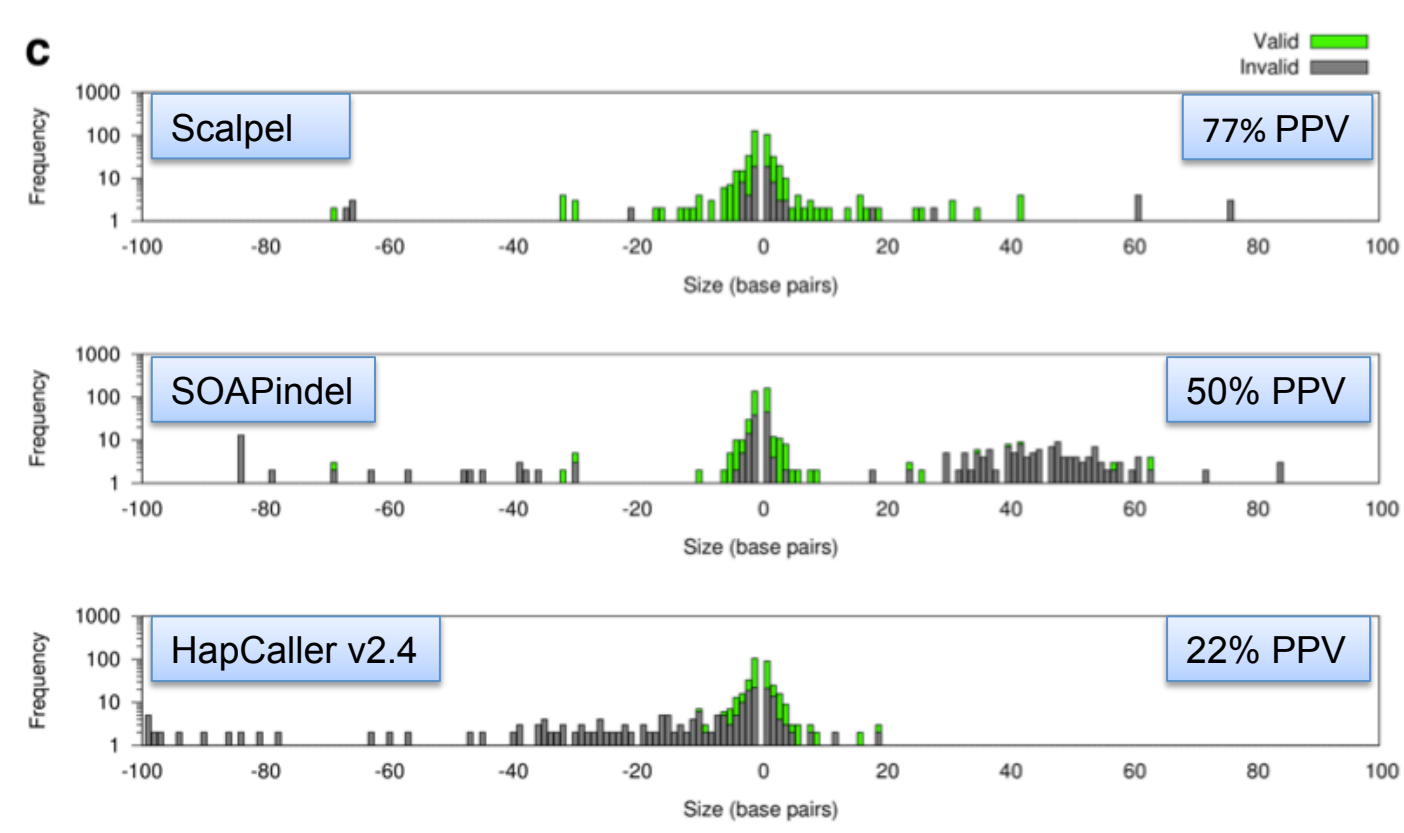
Accurate detection of de novo and transmitted INDELs within exome-capture data using micro-assembly

Giuseppe Narzisi, Jason A ORawe, Ivan Iossifov, Han Fang, Yoon-ha Lee, Zihua Wang, Yiyang Wu, Gholson J Lyon, Michael Wigler, Michael C Schatz

doi: 10.1101/001370

Narzisi et.al (Accepted in Nat. Methods)

Developing the best INDEL caller, with a large validation of 1400 INDELs



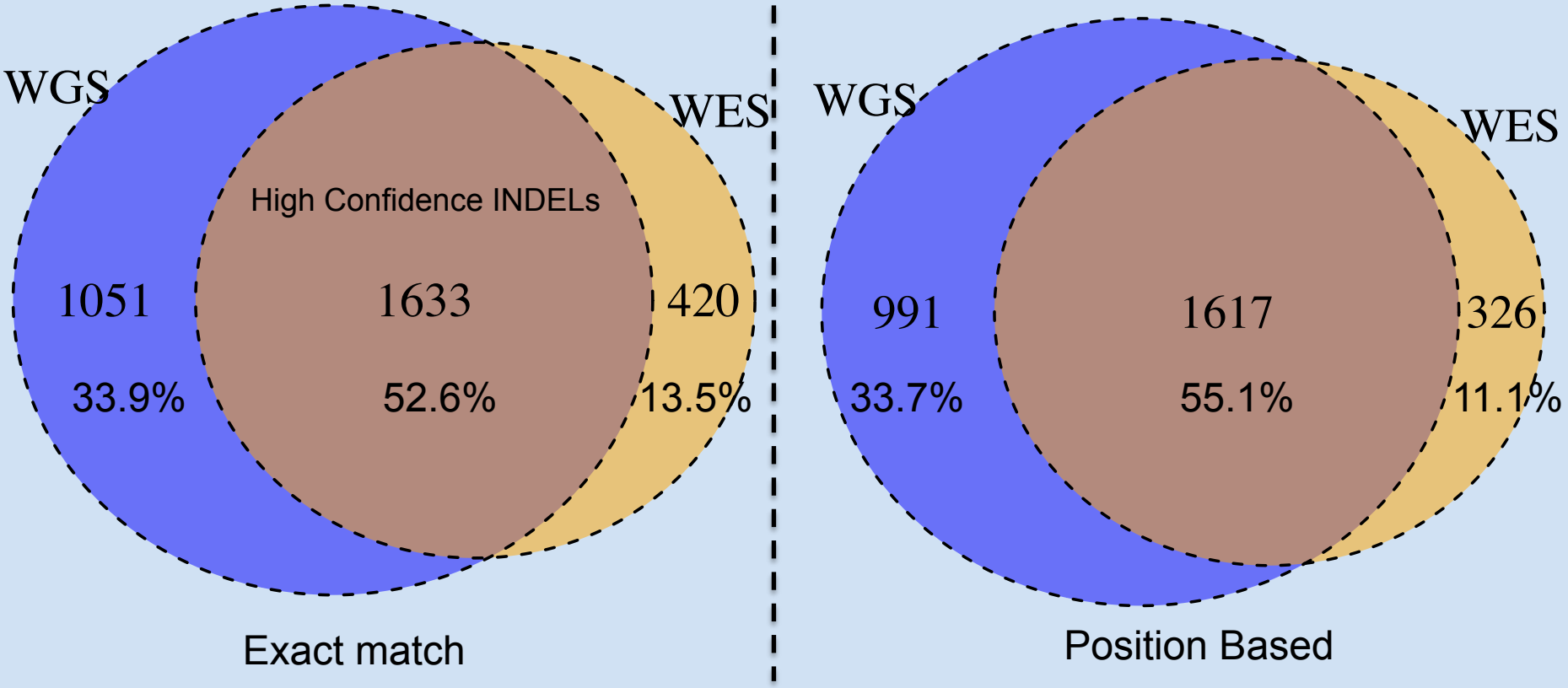
There are recent improvements with GATK v3.0 with 45% PPV, but Scalpel still out-performs this.

Narzisi et.al (Accepted in Nat. Methods)

Extending Scalpel with comparisons of WGS & WES data

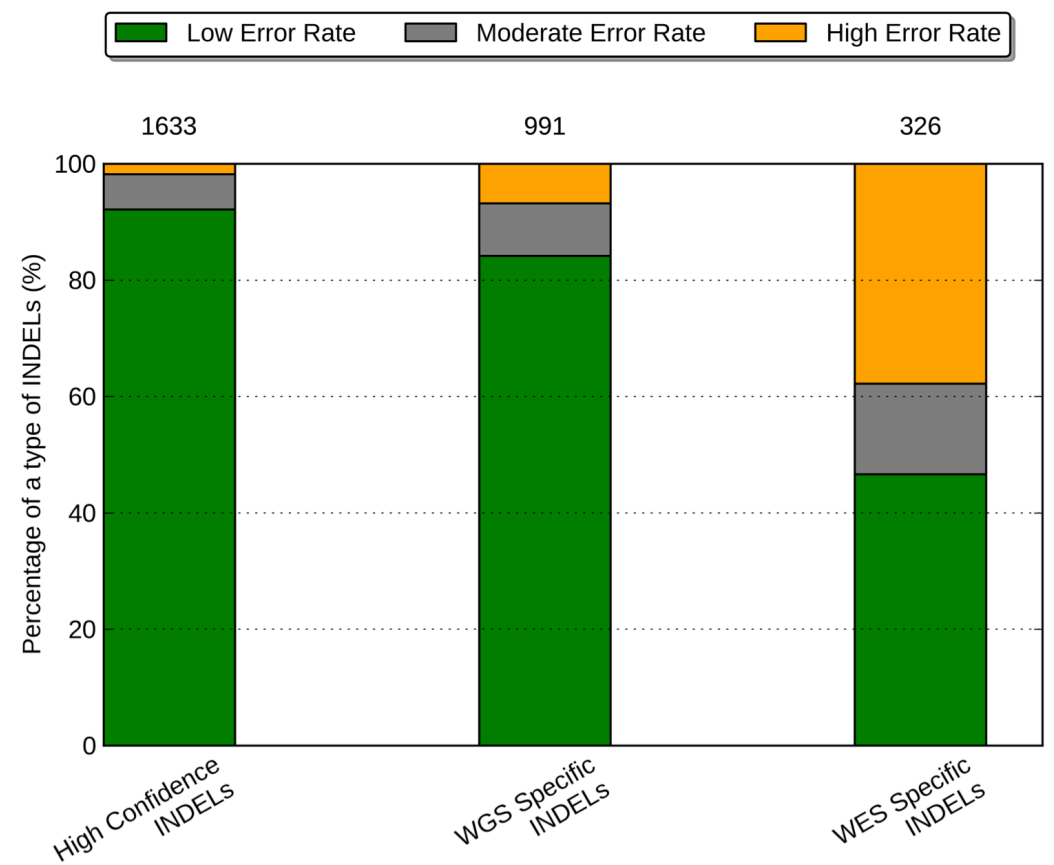
- WGS and WES were performed on 8 samples.
- Illumina HiSeq 2000 platform, paired-end 100 bp reads.
- Exome Capture Kit: NimbleGen SeqCap EZ Exome v2.0 capture reagent, representing 36.0 Mb (approximately 300,000 exons) of the human genome (hg19 build).
- WGS: Mean coverage= ~70x, ~95% > 20x
- WES: Mean coverage= ~320x, ~75% > 20x
- PCR duplicates were removed from the alignment.
- Inspected 25bp upstream and downstream around the loci of interest.

Mean concordance (8 samples) between WGS and WES data.



If keeping only regions in both data by requiring at least 1 read, the mean concordance rates increased to 62.1% (exact match) and 65.6% (positions based), respectively.

WGS yielded more “higher quality” INDELs, relative to WES.



Classification of call sets with previous validation data:

Low Error Rate: Coverage(alternative allele) >10 reads or χ^2 <4

High Error Rate: Coverage(alternative allele) <10 reads and χ^2 >10.84

Moderate Error Rate: Do not fall into the above two categories.

Note: The number on top of a category represents the mean number of INDELs in that category.

Previous works tried to understand coverage requirement for SNP calling. But how deep is deep enough for INDEL calling?

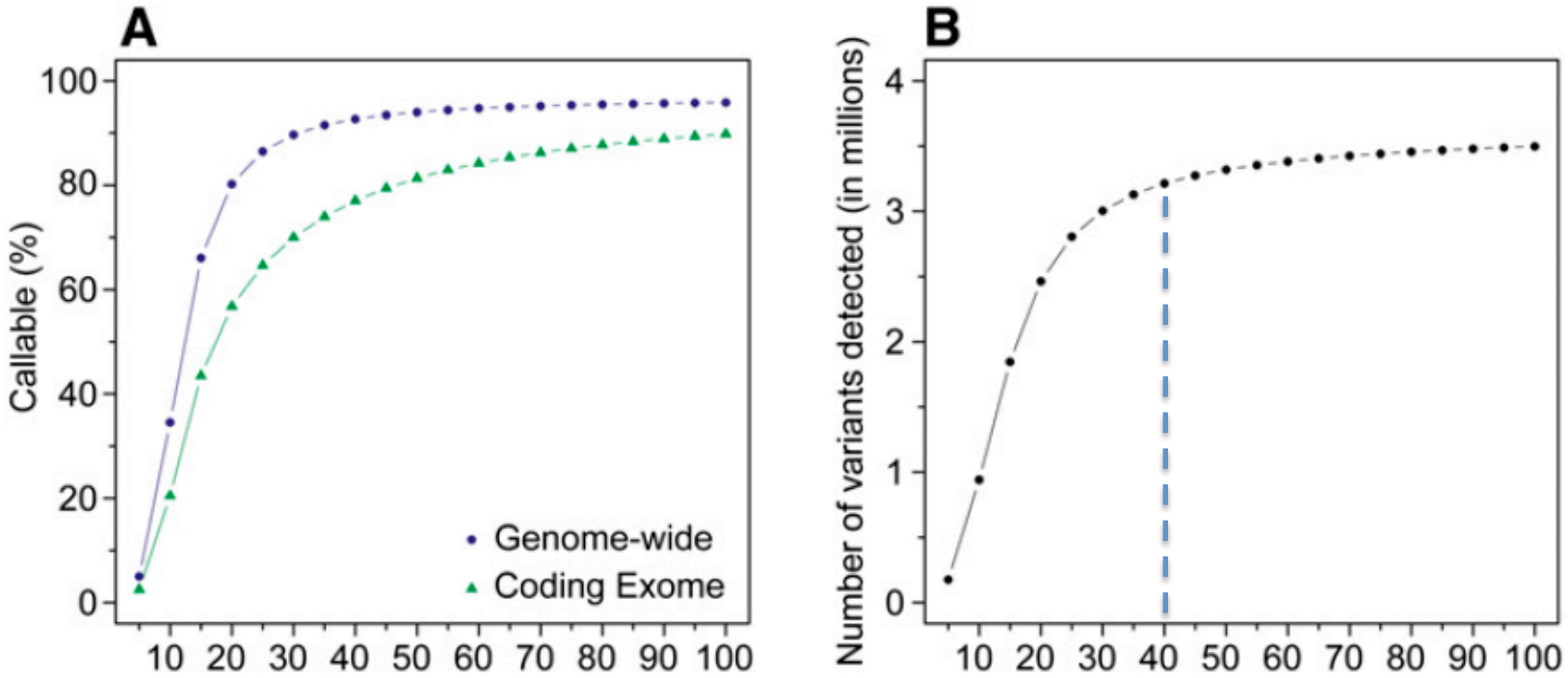
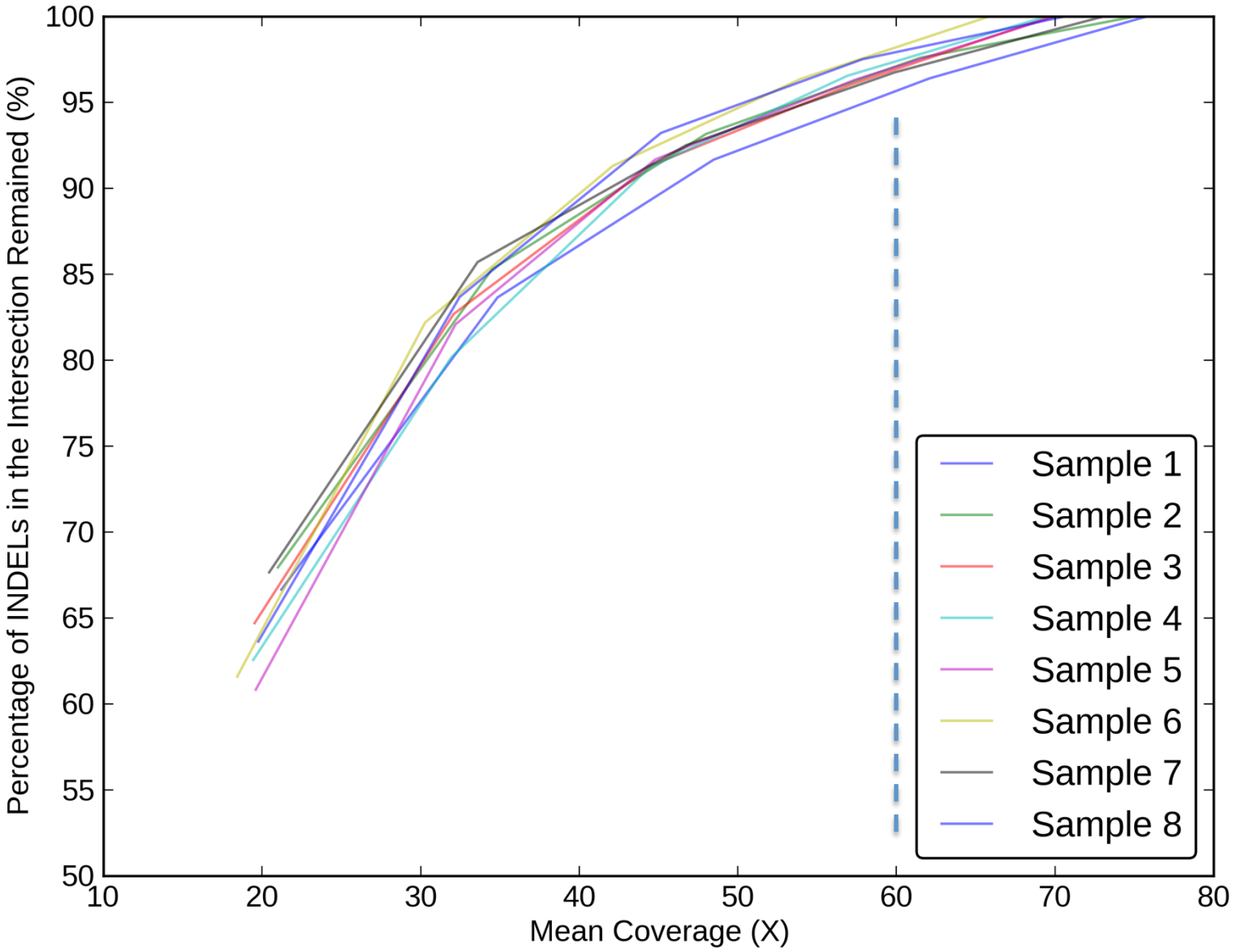


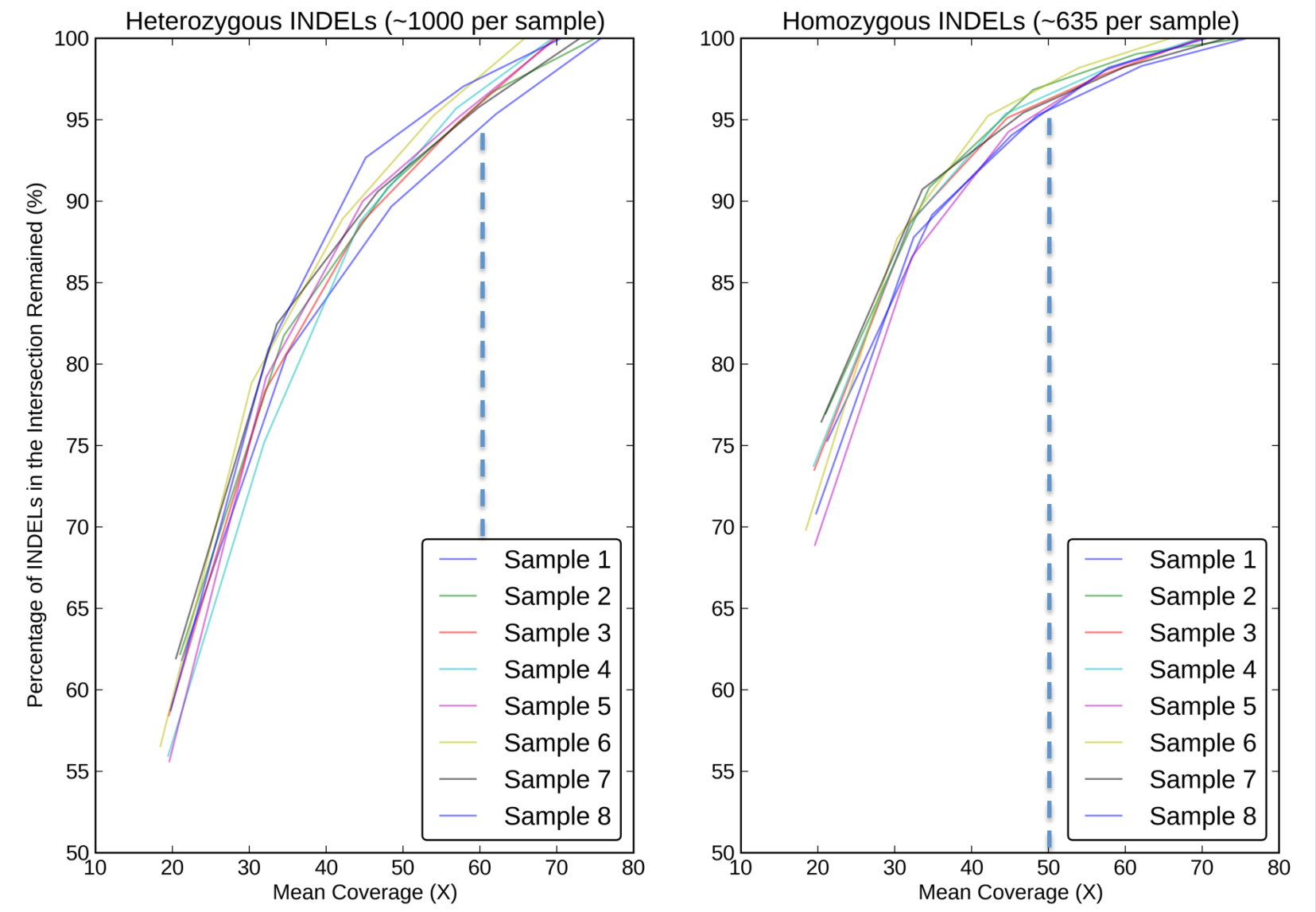
Figure 5. Genotype calling as a function of average mapped depth. The *x*-axes represent the average mapped depth of each data set, and the *y*-axes represent the proportion of the whole genome (dark blue circles) and coding exome (green triangles) that is callable (*A*), the number of SNVs detected (*B*), the proportion of Illumina BeadChip positions callable (*C*), and the concordance rates with the BeadChip calls (*D*).

Margulies et.al (2011)

Recommend mean coverage of 60X for personal genome sequencing to achieve high accuracy INDEL detection



Detection of heterozygous INDELs requires higher coverage; reaffirm the recommendation of 60X mean coverage



Limits of our current technology & knowledge

Analytic Validity

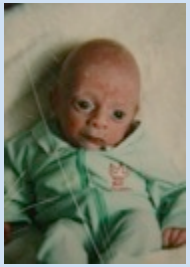
- Sequencing “clinical-grade genomes”

- Bioinformatics analysis

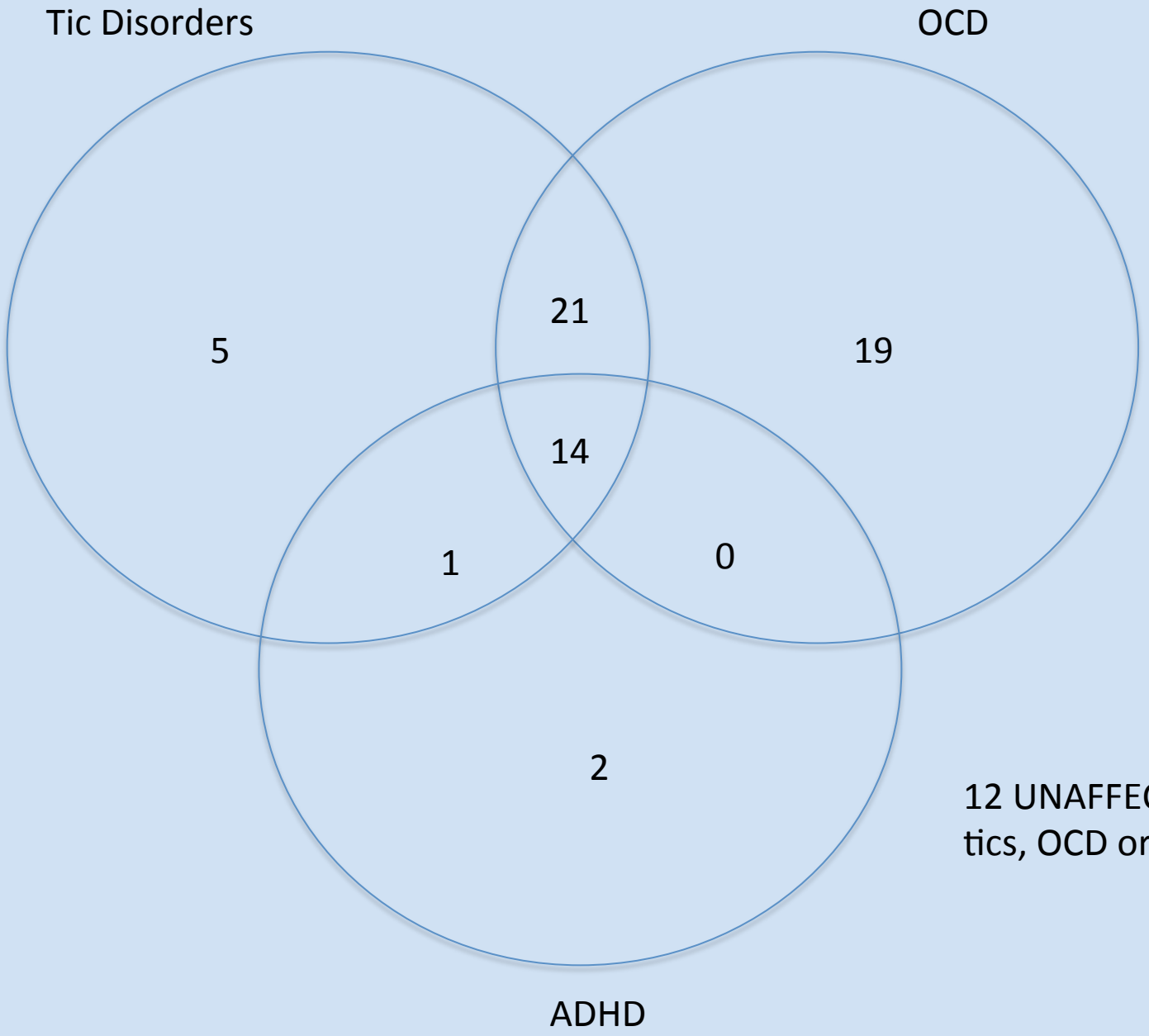
Clinical Validity

- **Genetic architecture of illness**

One family in Utah with a very rare disease.

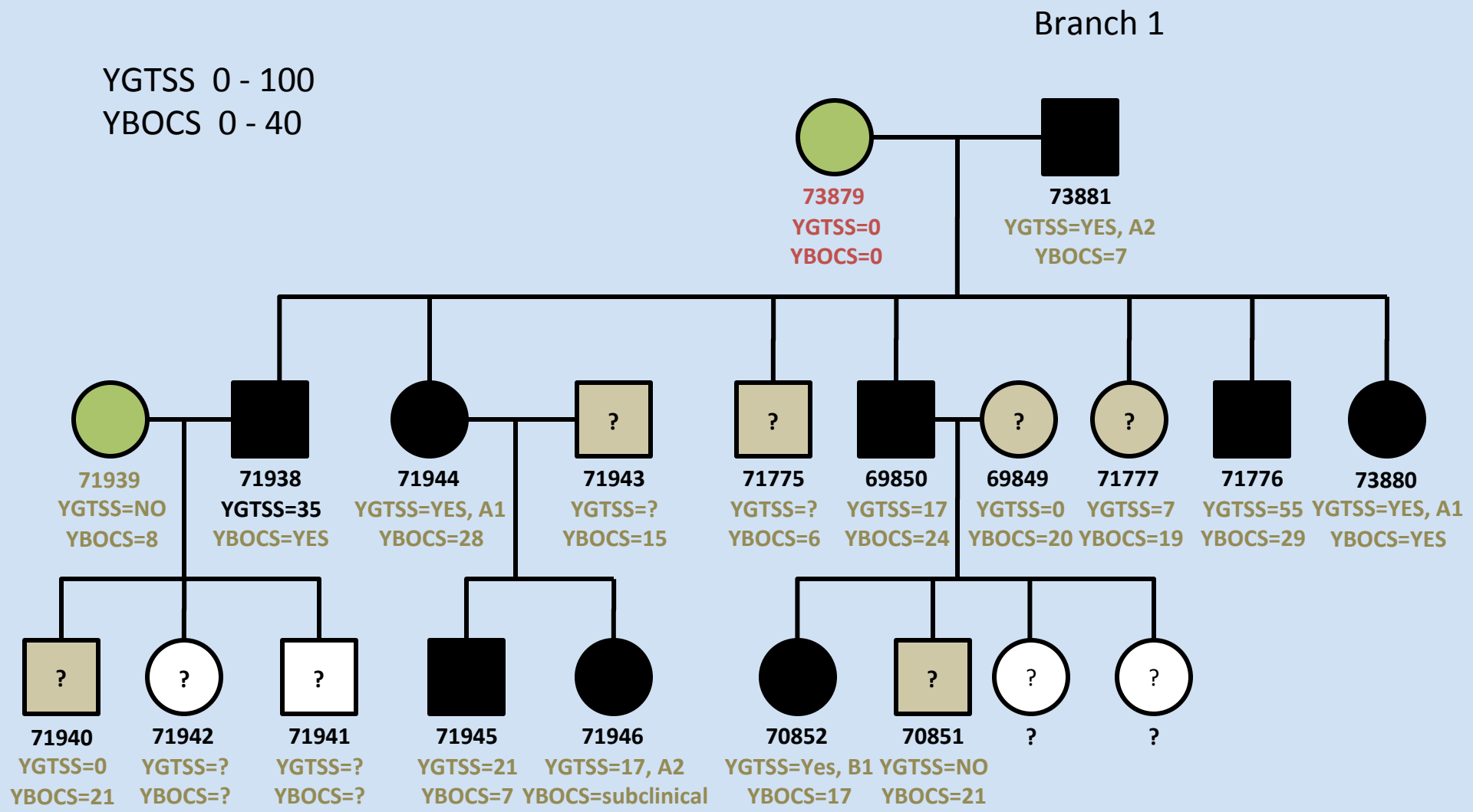


A Large Pedigree in Utah with mental illness



12 UNAFFECTED with tics, OCD or ADHD

Phenotyping of just one branch in this pedigree



Expression Issues

- We do not really know the expression of pretty much ALL mutations in **humans**, as we have not systematically sequenced or karyotyped any genetic alteration in **Thousands to Millions** of **randomly** selected people.

Molecular Genetics & Genomic Medicine

Open Access

ORIGINAL ARTICLE

Disease variants in genomes of 44 centenarians

Yun Freudenberg-Hua[1,2], Jan Freudenberg[3], Vladimir Vacic[4], Avinash Abhyankar[4], Anne-Katrin Emde[4], Danny Ben-Avraham[5], Nir Barzilai[5], Dayna Oschwald[4], Erika Christen[1], Jeremy Koppel[1,2], Blaine Greenwald[2], Robert B. Darnell[4,6], Soren Germer[4], Gil Atzmon[5] & Peter Davies[1]

[1]The Litwin-Zucker Research Center for the Study of Alzheimer's Disease and Memory Disorders, The Feinstein Institute for Medical Research, North Shore-LIJ, Manhasset, New York 11030

[2]Division of Geriatric Psychiatry, Zucker Hillside Hospital, North Shore-LIJ, Glen Oaks, New York 11040

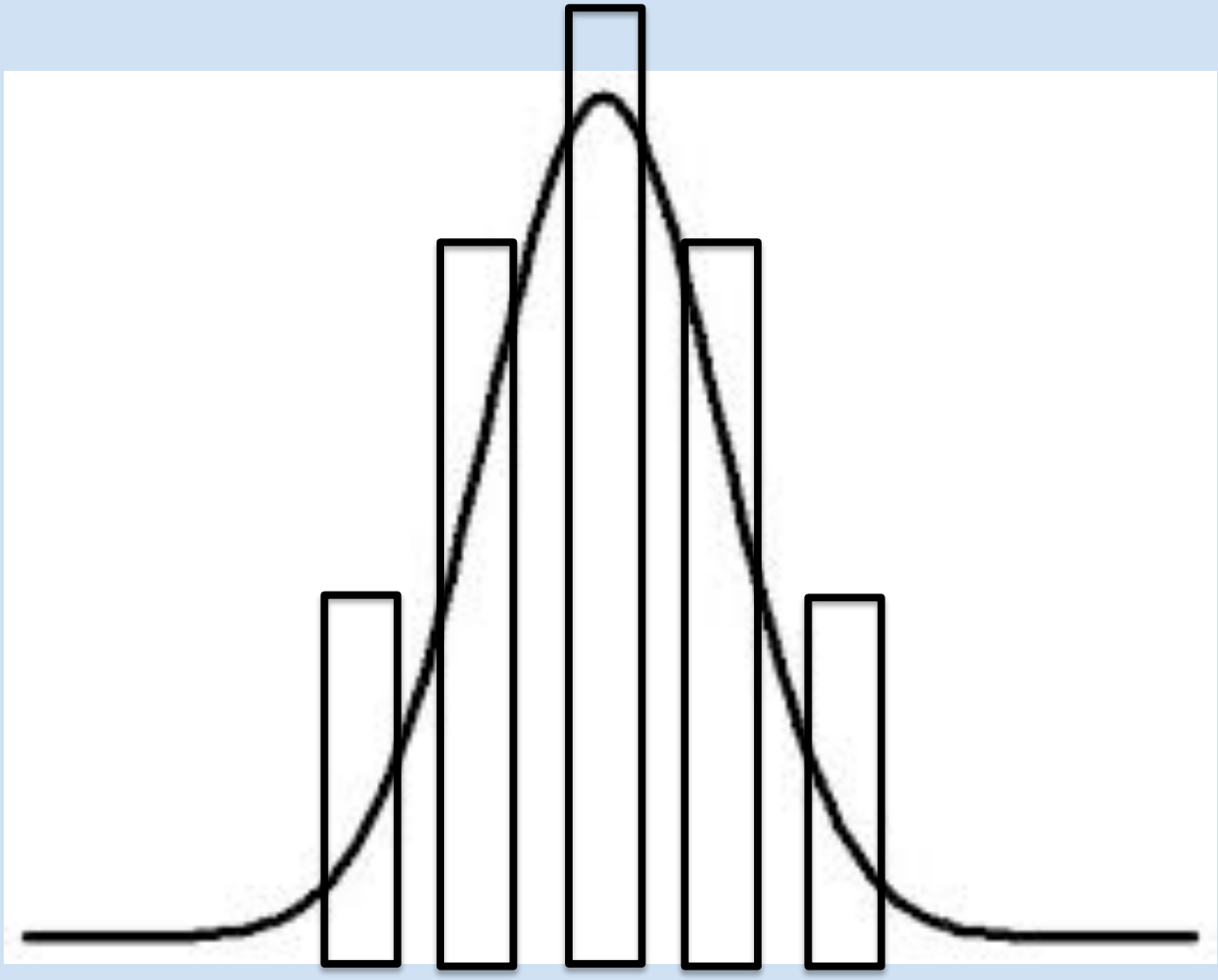
[3]Robert S. Boas Center for Genomics and Human Genetics, The Feinstein Institute for Medical Research, North Shore-LIJ, Manhasset, New York 11030

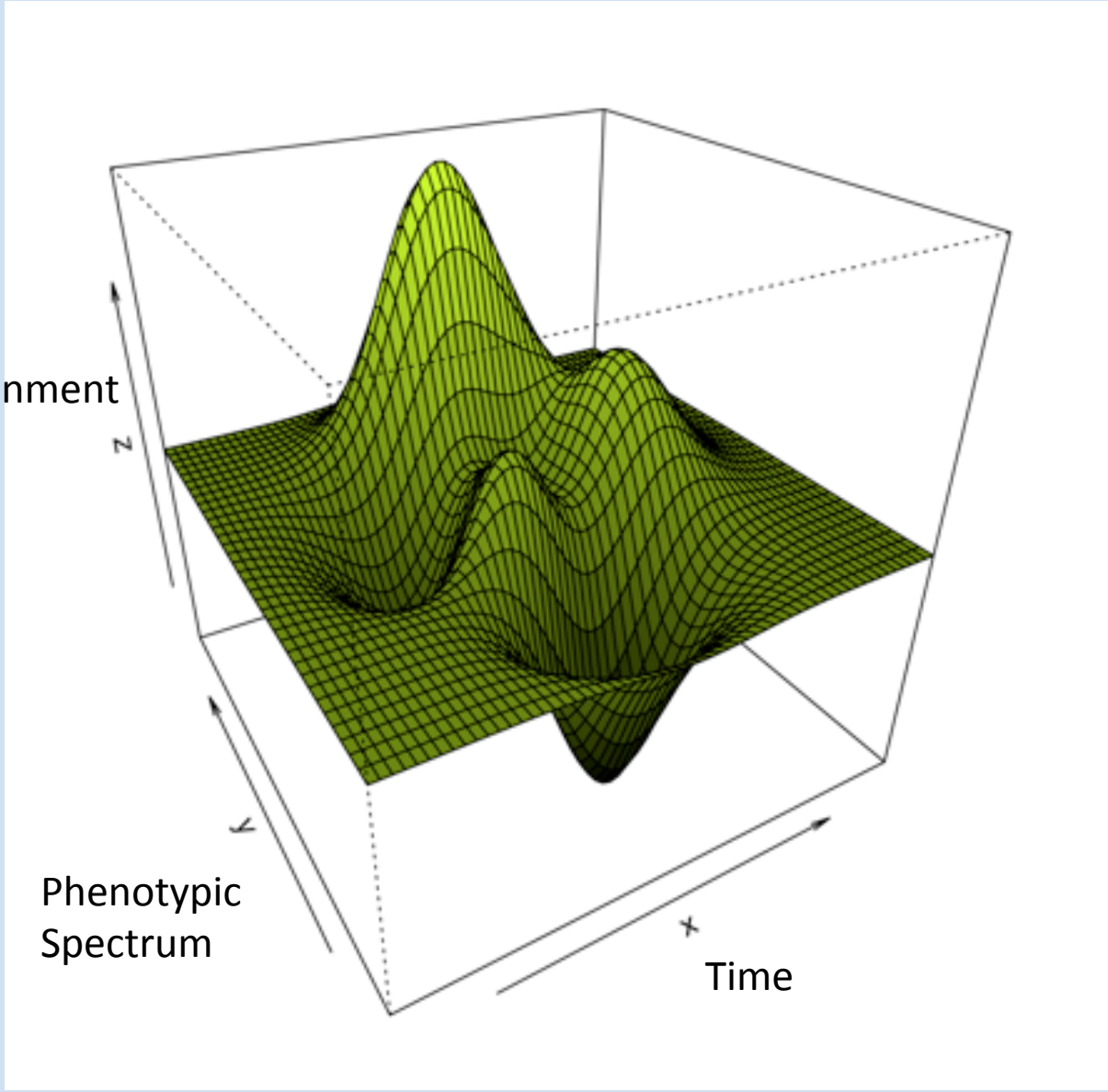
[4]New York Genome Center, 101 Avenue of the Americas, New York, New York 10013

[5]Institute for Aging Research Departments of Medicine and Genetics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461

[6]Department of Molecular Neuro-Oncology, Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10065

Categorical Thinking Misses Complexity





A conceptual model of genotype-phenotype correlations. The *y* plane represents a phenotypic spectrum, the *x* plane represents the canalized progression of development through time, and the *z* plane represents environmental fluctuations.

bioRxiv beta

THE PREPRINT SERVER FOR BIOLOGY

Clinical genetics of neurodevelopmental disorders

Gholson J Lyon and Jason O'Rawe

bioRxiv posted online November 18, 2013
Access the most recent version at doi:10.1101/000687

"There are ~12 billion nucleotides in every cell of the human body, and there are ~25-100 trillion cells in each human body. Given somatic mosaicism, epigenetic changes and environmental differences, no two human beings are the same, particularly as there are only ~7 billion people on the planet".

False Positives in the Literature

REPORT

XLID-Causing Mutations and Associated Genes Challenged in Light of Data From Large-Scale Human Exome Sequencing

Amélie Piton,[1,2,4,*] Claire Redin,[1,2,4] and Jean-Louis Mandel[1,2,3,*]

Because of the unbalanced sex ratio (1.3–1.4 to 1) observed in intellectual disability (ID) and the identification of large ID-affected families showing X-linked segregation, much attention has been focused on the genetics of X-linked ID (XLID). Mutations causing monogenic XLID have now been reported in over 100 genes, most of which are commonly included in XLID diagnostic gene panels. Nonetheless, the boundary between true mutations and rare non-disease-causing variants often remains elusive. The sequencing of a large number of control X chromosomes, required for avoiding false-positive results, was not systematically possible in the past. Such information is now available thanks to large-scale sequencing projects such as the National Heart, Lung, and Blood (NHLBI) Exome Sequencing Project, which provides variation information on 10,563 X chromosomes from the general population. We used this NHLBI cohort to systematically reassess the implication of 106 genes proposed to be involved in monogenic forms of XLID. We particularly question the implication in XLID of ten of them (*AGTR2*, *MAGT1, ZNF674, SRPX2, ATP6AP2, ARHGEF6, NXF5, ZCCHC12*, *ZNF41*, and *ZNF81*), in which truncating variants or previously published mutations are observed at a relatively high frequency within this cohort. We also highlight 15 other genes (*CCDC22*, *CLIC2*, *CNKSR2*, *FRMPD4*, *HCFC1*, *IGBP1*, *KIAA2022*, *KLF8*, *MAOA, NAA10*, *NLGN3*, *RPL10, SHROOM4*, *ZDHHC15*, and *ZNF261*) for which replication studies are warranted. We propose that similar reassessment of reported mutations (and genes) with the use of data from large-scale human exome sequencing would be relevant for a wide range of other genetic diseases.

Bring clinical standards to human-genetics research

Study protocols need to be rigorous, because more than science is at stake. Sometimes participants' lives depend on the results, writes **Gholson J. Lyon**.

Clinical Validity?

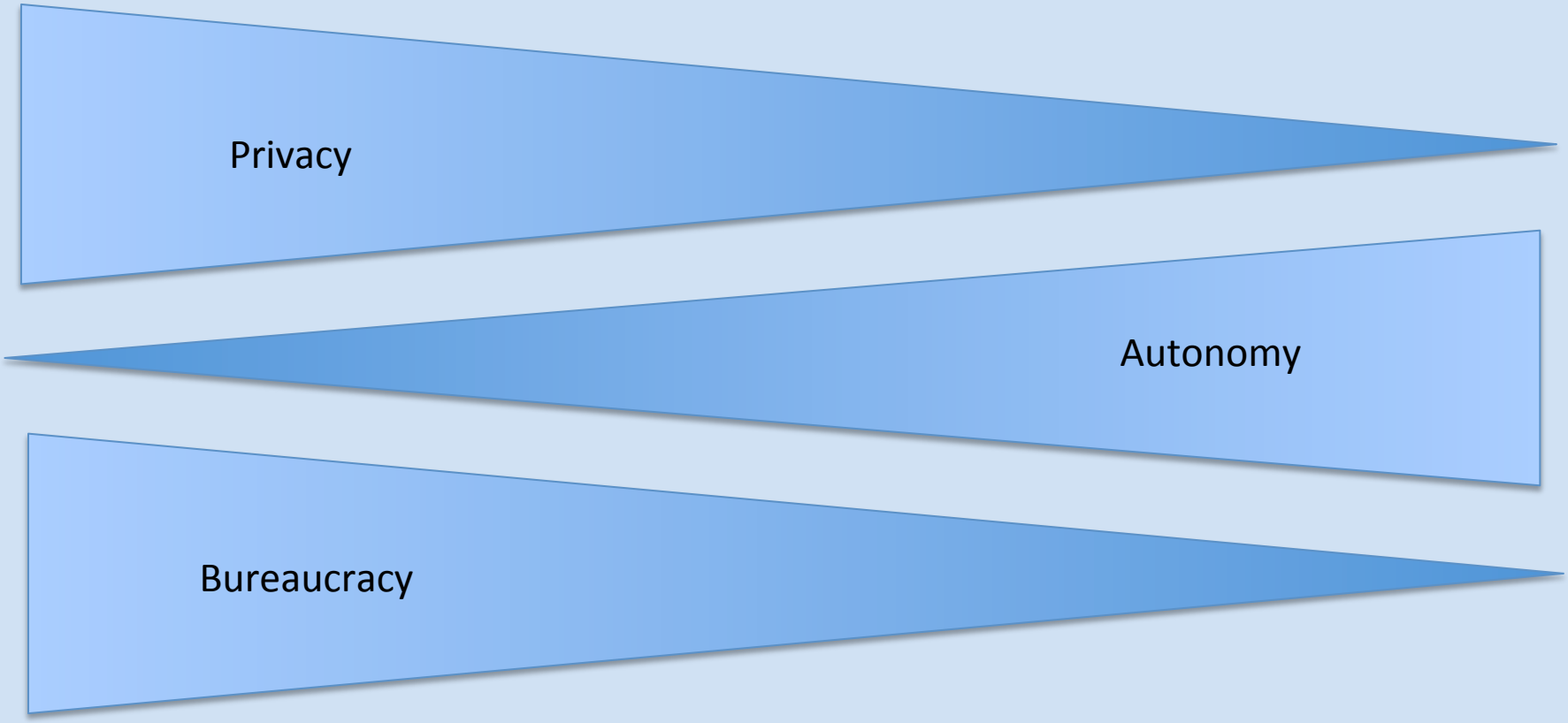
This is SO complex that the only solid way forward is with a “networking of science” model, i.e. online database with genotype and phenotype longitudinally tracked for thousands of volunteer families.

PatientsLikeMe

Major barriers to the widespread implementation of genomic medicine in the clinic.

- Limits of our current technology & knowledge
- Lack of public education
- Lack of physician knowledge about genetics
- Apathy on the part of the populace in terms of preventive efforts
- Reluctance of insurance companies & governments to pay for genetic testing
- Focus in our society on treatment, not on early diagnosis and prevention
- Privacy concerns

Autonomy vs. Privacy vs. Bureaucracy



PRIVACY and PROGRESS in Whole Genome Sequencing

Presidential Commission
for the Study of Bioethical Issues

October 2012

Published online 4 October 2011 | *Nature* **478**, 17 (2011) |
doi:10.1038/478017a

News

Secrets of the human genome disclosed

Meeting debates ethics of revealing genetic findings.

Erika Check Hayden

NEW YORK

Should people be told about any nasty surprises that scientists discover in their DNA during research projects?

The question is becoming increasingly pertinent, as thousands of people sign up for studies in which their genomes will be sequenced. But, at present, federal laws in the United States prohibit researchers from telling patients about mutations that might affect them or their families unless a certified clinical lab has confirmed the results — something that is not done in most research projects. This means that patients often do not learn about their mutations until the studies are finally published, a restriction that is meant to ensure they are not misinformed by incomplete research.

> **"These disclosures have societal implications that need to be considered."**

The ethical dilemmas became all too real last year for geneticist Gholson Lyon, a geneticist at the Utah Foundation for Biomedical Research in Salt Lake City. He was studying an extended family in which some of the boys had been born with a constellation of symptoms, including thick, wrinkly skin, and who ultimately died of cardiac disease before their first birthdays. By November 2010, Lyon had convincing evidence that a genetic mutation was causing the disease. That's when he learned that one of the women in the family was four months pregnant with a boy.

Policy and Governance

"If you sequence people's exomes you're going to find stuff," said Gholson Lyon, a physician and researcher previously at the University of Utah, now at Cold Spring Harbor Laboratory.

As part of his research, Dr. Lyon worked with a family in Ogden, Utah. Over two generations, four boys had died from an unknown disease with a distinct combination of symptoms—an aged appearance, facial abnormalities, and developmental delay. Dr. Lyon sought to identify the genetic cause of this disease, and collected blood samples from 12 family members who had signed consent forms. The family members understood these forms to mean that they would have access to their results.

Dr. Lyon has become an outspoken advocate for conducting whole genome sequencing in laboratories that satisfy the federal standards so that researchers can return results to participants, if appropriate. Dr. Lyon wants clear guidance for laboratories conducting genetic research and clear language in consent forms that clarifies the results that participants should expect to have returned from the researchers.

Recommendation 4.1

Funders of whole genome sequencing research, relevant clinical entities, and the commercial sector should facilitate explicit exchange of information between genomic researchers and clinicians, while maintaining robust data protection safeguards, so that whole genome sequence and health data can be shared to advance genomic medicine.

Performing all whole genome sequencing in CLIA-approved laboratories would remove one of the barriers to data sharing. It would help ensure that whole genome sequencing generates high-quality data that clinicians and researchers can use to draw clinically relevant conclusions. It would also ensure that individuals who obtain their whole genome sequence data could share them more confidently in patient-driven research initiatives, producing more meaningful data. That said, current sequencing technologies and those in development are diverse and evolving, and standardization is a substantial challenge. Ongoing efforts, such as those by the Standardization of Clinical Testing working group are critical to achieving standards for ensuring the reliability of whole genome sequencing results, and facilitating the exchange and use of these data.[216]

Contents lists available at SciVerse ScienceDirect

Applied & Translational Genomics

journal homepage: www.elsevier.com/locate/atg

Practical, ethical and regulatory considerations for the evolving medical and research genomics landscape

Gholson J. Lyon [a,b,*], Jeremy P. Segal [c,**]

[a] Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, NY, United States
[b] Utah Foundation for Biomedical Research, Salt Lake City, UT, United States
[c] New York Genome Center, New York City, NY, United States

Table 1
Processes involved in a CLIA-certified genetic test.

Preanalytic system
1) Test request and specimen collection criteria
2) Specimen submission, handling and referral procedures
3) Preanalytic systems assessment
Analytic system
1) A detailed step-by-step procedure manual
2) Test systems, equipment, instruments, reagents, materials and supplies
3) Establishment and verification of performance specifications
4) Maintenance and function checks
5) Calibration and calibration verification procedures
6) Control procedures, test records, and corrective actions
7) Analytic systems assessment
Post-analytic system
1) Test report, including (among other things):
a) interpretation
b) reference ranges and normal values
2) Post-analytic systems assessment

1. Sample Collection and handling

2. Sequencing/Analytics

3. Interpretation

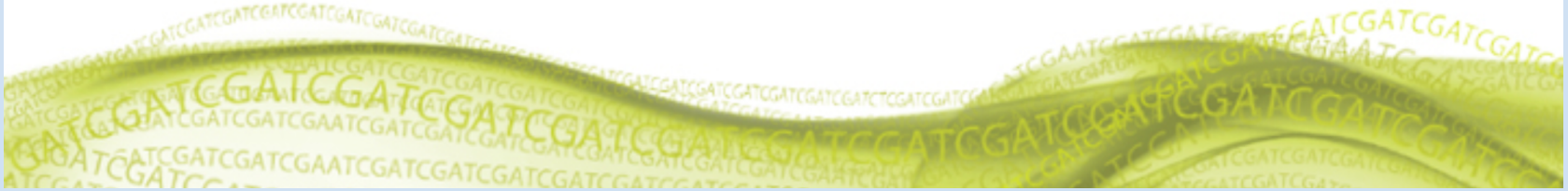
Individual Genome Sequencing Service

Available from Illumina's CLIA-certified laboratory.

"This laboratory test was developed, and its performance characteristics were determined by the Illumina Clinical Services Laboratory (CLIA-certified, CAP-accredited). Consistent with laboratory-developed tests, it has not been cleared or approved by the U.S. Food and Drug Administration. If you have any questions or concerns about what you might learn through your genome sequence information, you should contact your doctor or a genetic counselor. Please note that Illumina does not accept orders for Individual Genome Sequencing services from Florida and New York."

Understand Your Genome Symposium

During this two-day educational event, industry experts will discuss the clinical implementation of whole-genome next-generation sequencing (NGS) technology.



illumina®

Ordering Physician:
Gholson Lyon, MD
Steinmann Institute
10 West Broadway, Suite #820
Salt Lake City, UT 84101

Individual Genome Sequence Results

Clinical Report

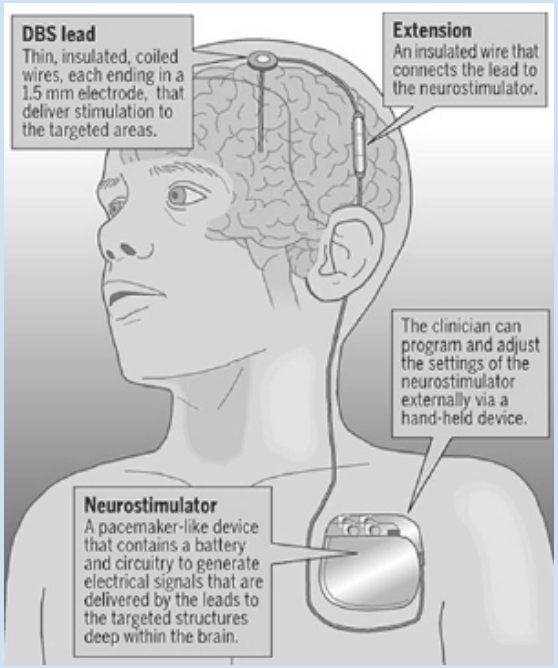
www.everygenome.com
CLIA#: 05D1092911

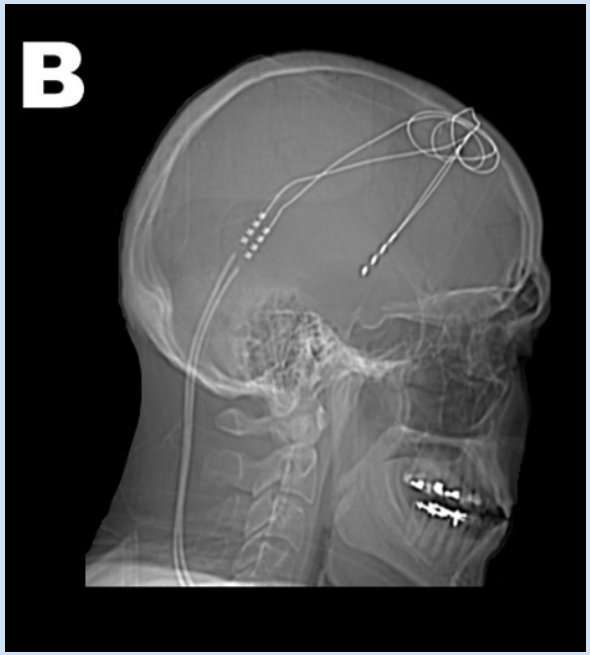
PeerJ

Integrating precision medicine in the study and clinical treatment of a severely mentally ill person

Jason A. O'Rawe[1,2], Han Fang[1,2], Shawn Rynearson[3], Reid Robison[4], Edward S. Kiruluta[5], Gerald Higgins[6], Karen Eilbeck[3], Martin G. Reese[5] and Gholson J. Lyon[1,2,4]

[1] Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, NY, USA
[2] Stony Brook University, Stony Brook, NY, USA
[3] Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA
[4] Utah Foundation for Biomedical Research, Salt Lake City, UT, USA
[5] Omicia Inc., Emeryville, CA, USA
[6] AssureRx Health, Inc., Mason, OH, USA





Submitted 12 June 2013
Accepted 16 September 2013
Published 3 October 2013

Corresponding author
Gholson J. Lyon,
GholsonJLyon@gmail.com

Academic editor
Paul Appelbaum

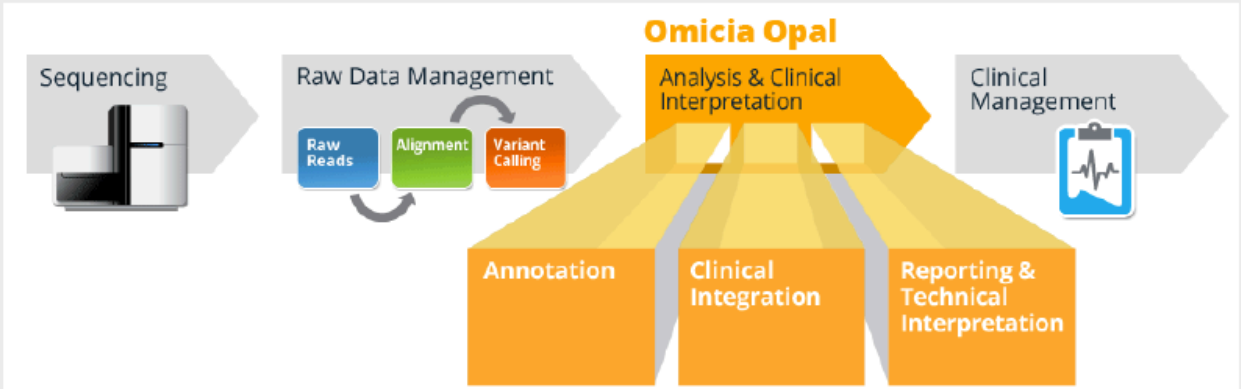
Additional Information and Declarations can be found on page 18

DOI **10.7717/peerj.177**

Commercial analysis platforms for genomic data

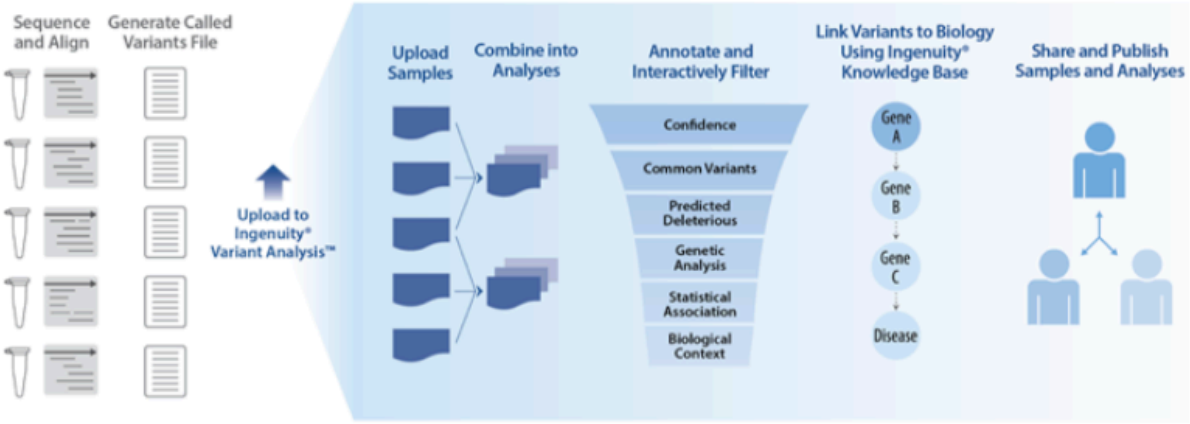
Opal adds clinical context for genomic data

Omicia is unlocking individualized medicine by translating data derived from whole-genome sequencing into actionable information for researchers and clinicians.



Omicia
Unlocking Individualized Medicine

Identify causal variants from human sequencing data in just hours



BIOLOGICAL INTERPRETATION OF HUMAN WHOLE GENOME, EXOME, AND TARGETED PANEL SAMPLES

SNP & VARIATION SUITE 7

SNP & Variation Suite 7 is an integrated collection of user-friendly, yet powerful analytic tools for managing, analyzing, and visualizing multifaceted genomic and phenotypic data. SVS was created specifically to empower biologists and other researchers to easily perform complex analyses and visualizations, eliminating the need to rely exclusively on bioinformatics experts or cobble together difficult to use, incompatible freeware. With SVS you can focus on your research instead of learning to be a programmer or waiting in line for bioinformaticians.

SNP CNV DNA-Seq RNA-Seq Power PBAT Core Plus Viewer

No rare variants or CNVs with high biological effect as related to mental illness.

Here are just 3 of dozens of common SNVs in this person that have been implicated in the literature as predisposing to mental illness.

Table 1 A summary of three clinically relevant alleles found in the sequencing results of MA. Variations in MTHFR, BDNF, and ChAT were found to be of potential clinical relevance for this person as they are all implicated in contributing to the susceptibility and development of many neuropsychiatric disorders that resemble those present within MA. A brief summary of the characteristics of each variation is shown, including the gene name, genomic coordinates, amino acid change, zygosity, variation type, estimated population frequency and putative clinical significance.

Gene name	Genomic coordinates	Amino acid change	Zygosity	Variation type	Population frequency	Clinical significance
MTHFR	chr1: 11854476	Glu > Ala	heterozygous	non-synon	T:77% G:23%	Susceptibility to psychoses, schizophrenia occlusive vascular disease, neural tube defects, colon cancer, acute leukemia, and methylenetetra-hydrofolate reductase deficiency
BDNF	chr11: 27679916	Val > Met	heterozygous	non-synon	C:77% T:23%	Susceptibility to OCD, psychosis, and diminished response to exposure therapy
CHAT	chr10: 50824117	Asp > Asn	heterozygous	non-synon	G:85% A:15%	Susceptibility to schizophrenia and other psy-chopathological disorders.

Table 3 Description of the 22 genome-wide significant loci in the combined analysis

Chromosomal region	*P* value	Previous association[a]	Candidate gene in relation to index SNP[b]	Other genes in genomic region defined by LD[c]	eQTL[d]	Disease associations[e]
Chr. 6: 31,596,138–32,813,768	9.14×10^{-14}	SCZ	*HLA-DRB9*	MHC class II, many other genes, lincRNA	Many	Many
Chr. 10: 104,487,871–105,245,420	3.68×10^{-13}	SCZ	*C10orf32-AS3MT*	*CALHM1*, *CALHM2*, *CALHM3*, *CNNM2*, *CYP17A1*, *INA*, *MIR1307*, *NT5C2*, *PCGF6*, *PDCD11*, *SFXN2*, *ST13P13*, *TAF5*, *USMG5*, *WBP1L*	*ACTR1A*, *ARL3*, *AS3MT*, *C10orf32*, *C10orf78*, *NT5C2*, *TMEM180*, *TRIM8*, *WBP1L*	GWAS: blood pressure, CAD, aneurysm
Chr. 7: 1,827,717–2,346,115	5.93×10^{-13}	No	*MAD1L1*	*FTSJ2*, *NUDT1*, *SNX8*	*C7orf27*, *FTSJ2*, *MAD1L1*, *NUDT1*	
Chr. 1: 98,141,112–98,664,991	1.72×10^{-12}	SCZ	(*MIR137*, 37 kb)	*DPYD*, lincRNA	***DPYD***	*DPYD*: mental retardation
Chr. 12: 2,285,731–2,440,464	5.22×10^{-12}	SCZ, BPD	*CACNA1C*	–	No data	*CACNA1C*: autism, Timothy syndrome, Brugada syndrome 3
Chr. 10: 18,601,928–18,934,390	1.27×10^{-10}	5 disorders	*CACNB2*	*NSUN6*	No data	*CACNB2*: Brugada syndrome 4; GWAS: blood pressure
Chr. 8: 143,297,312–143,410,423	2.19×10^{-10}	No	*TSNARE1*	–	No data	
Chr. 1: 73,275,828–74,099,273	3.64×10^{-10}	No	(x10NST00000415686.1, 4 kb)	lincRNA	No data	
Chr. 11: 130,706,918–130,894,976	1.83×10^{-9}	No	(*SNX19*, 31 kb)	lincRNA	***SNX19***	
Chr. 5: 151,888,959–152,835,304	2.65×10^{-9}	No	ENST00000503048.1	lincRNA (*GRIA1*)	No data	
Chr. 5: 152,505,453–152,707,306	4.12×10^{-8}	No				
Chr. 19: 19,354,937–19,744,079	3.44×10^{-9}	BPD	(*MAU2*, 4 kb)	*CILP2*, *GATAD2A*, *GMIP*, *HAPLN4*, *LPAR2*, *MIR640*, *NCAN*, *NDUFA13*, *PBX4*, *SUGP1*, *TM6SF2*, *TSSK6*, *YJEFN3*	No data	GWAS: lipid levels

[a]Regions reported to meet genome-wide significance thresholds of association for schizophrenia (SCZ) or bipolar disorder (BPD). [b]The gene within which an index SNP is located is given. For intergenic index SNPs, the nearest gene is given in parentheses. [c]Other named genes in the genomic interval. [d]SNP-transcript associations with $q < 0.05$ in peripheral blood. eQTLs with the SNP with the strongest association are shown in bold. [e]Data from the NHGRI GWAS catalog[24], OMIM[43] and a compilation of genes related to autism[73] and mental retardation[43,74,75]. No data means no Affymetrix U219 probe sets or low expression in peripheral blood. The *CACNB2* association emerged when considering attention deficit/hyperactivity disorder (ADHD), autism, bipolar disorder, major depressive disorder and schizophrenia as affected[30]. CAD, coronary artery disease; HDL, high-density lipoprotein.

Indicates that M.A. is homozygous for the exact variant of genome significance

Indicates that M.A. is heterozygous for the exact variant of genome significance

Ripke et.al (2013) Nature Genetics

	Chr. 2: 37,422,072–37,592,628	6.78×10^{-9}	No	*QPCT*	*C2orf56*, *CEBPZ*, *PRKD3*, *SULT6B1* lincRNA	No eQTL	
	Chr. 5: 101,581,848–101,870,822	9.03×10^{-9}	No	*SLCO6A1*	lincRNA	No data	
	Chr. 3: 52,215,002–53,175,017	1.16×10^{-8}	SCZ, BPD	*ITIH3*	*ALAS1*, *ALDOAP1*, *BAP1*, *C3orf78*, *DNAH1*, *GLT8D1*, *GLYCTK*, *GNL3*, *ITIH1*, *ITIH4*, *MIR135A1*, *MIRLET7G*, *MUSTN1*, *NEK4*, *NISCH*, *NT5DC2*, *PBRM1*, *PHF7*, *PPM1M*, *RFT1*, *SEMA3G*, *SFMBT1*, *SPCS1*, *STAB1*, *TLR9*, *TMEM110*, *TNNC1*, *TWF2*, *WDR82*, lincRNA	No data (*ITIH1-ITIH3-ITIH4*)	*GLYCTK*: D-glyceric aciduria, mental retardation; *RTF1*: mental retardation; GWAS: adiponectin, height, waist-hip ratio
	Chr. 2: 145,139,727–145,214,607	1.19×10^{-8}	No	*ZEB2*	–	No eQTL	*ZEB2*: Mowat-Wilson syndrome, mental retardation
	Chr. 2: 200,628,118–201,293,421	1.21×10^{-8}	No	*FONG*	*C2orf47*, *C2orf69*, *SPATS2L*, *TYW5*, lincRNA	No data	GWAS: osteoporosis
	Chr. 18: 52,722,378–52,827,668	1.22×10^{-8}	No	(ENST00000565991.1, 21 kb)	lincRNA (*TCF4*)	No data	
	Chr. 2: 233,550,961–233,808,241	1.51×10^{-8}	No	*C2orf82*	*GIGYF2*, *KCNJ13*, *NGEF*	No data	
	Chr. 1: 243,593,066–244,025,999	1.80×10^{-8}	No	*AKT3*	*CEP170*	***AKT3***	
	Chr. 1: 243,418,063–243,627,135	2.53×10^{-8}	Yes	*SDCCAG8*		***SDCCAG8***	
	Chr. 12: 123,447,928–123,913,433	2.28×10^{-8}	No	*C12orf65*	*ABCB9*, *ARL6IP4*, *CDK2AP1*, *MIR4304*, *MPHOSPH9*, *OGFOD2*, *PITPNM2*, *RILPL2*, *SBNO1*, *SETD8*, lincRNA	*ARL6IP4*, *CDK2AP1*, *OGFOD2*, *SBNO1*	*C12orf65*: mental retardation; GWAS: HDL, height, head size
	Chr. 8: 89,188,454–89,761,163	3.33×10^{-8}	SCZ	Intergenic	*MMP16*, lincRNA	*MMP16*	
	Chr. 5: 60,484,179–60,843,706	3.78×10^{-8}	No	ENST00000506902.1	*ZSWIM6*, *C5orf43*, lincRNA	*C5orf43*, *ZSWIM6*	

Indicates that M.A. is homozygous for the exact variant of genome significance

Indicates that M.A. is heterozygous for the exact variant of genome significance

Ripke et.al (2013) Nature Genetics

Pharmacogenetics

- MA is homozygous for a p.Ile359Leu change in CYP2C9, and this variant has been linked to a reduction in the enzymatic activity of CYP2C9, a member of the cytochrome P450 superfamily of enzymes.
- Fluoxetine is commonly used in the treatment of OCD..
- CYP2C9 acts to convert fluoxetine to R-norfluoxetine, and so MA may not be able to adequately biotransform fluoxetine.
- It is notable that MA had no response to an 80 mg daily dose of fluoxetine.
- However, CYP2C9 does not play a rate-limiting role for other SSRIs or clomipramine

Clinical Validity with Worldwide Human Genetic Variation "database"?

PatientsLikeMe

100,000 British Genomes

Clinical Validity?

This is SO complex that the only solid way forward is with a “networking of science” model, i.e. online database with genotype and phenotype longitudinally tracked for thousands of volunteer families.

PatientsLikeMe

Learning Objectives

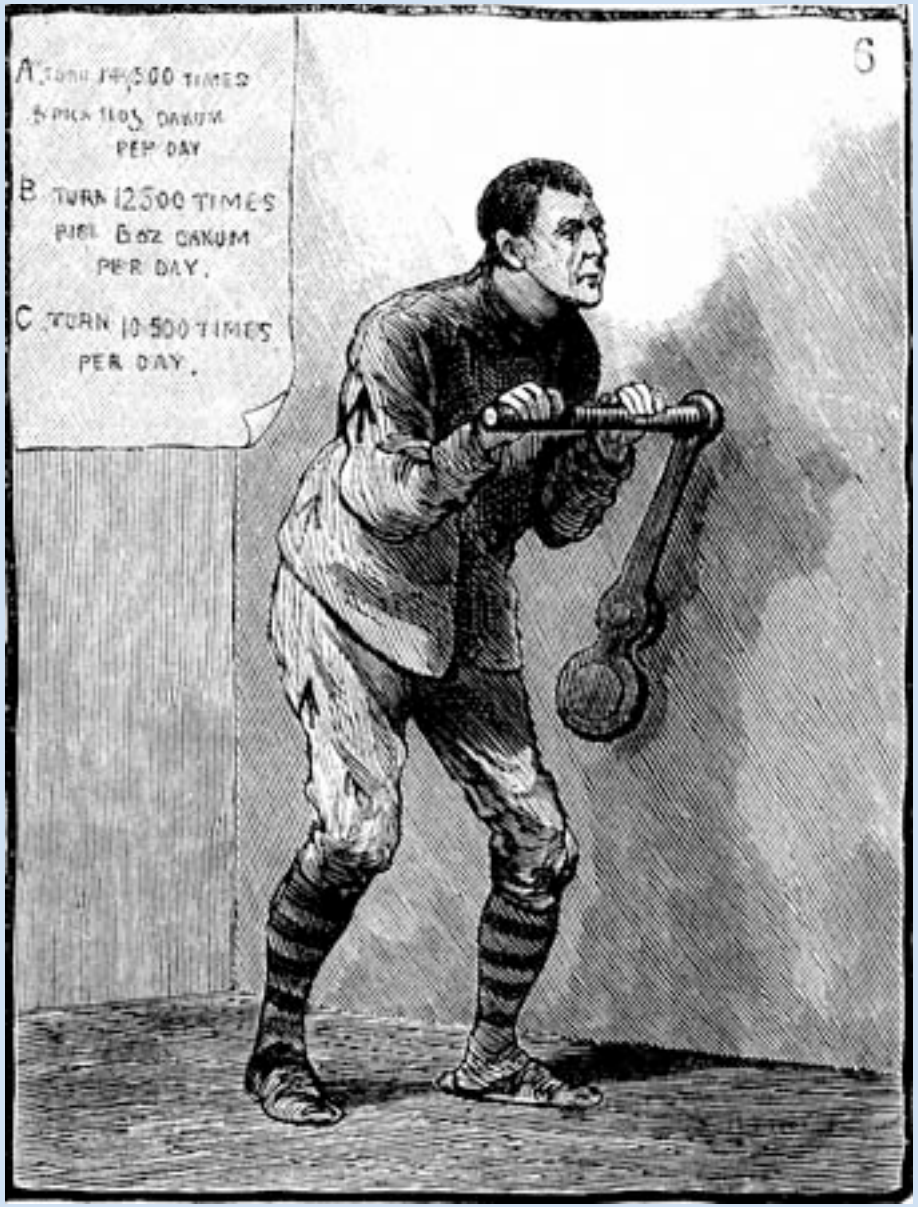
- The participants will be able to discuss the extraordinary amount of variable expressivity seen in neurodevelopmental disorders.

- The participants will plan ways in which to integrate genomic and phenotypic longitudinal data to prevent the development of certain illnesses.

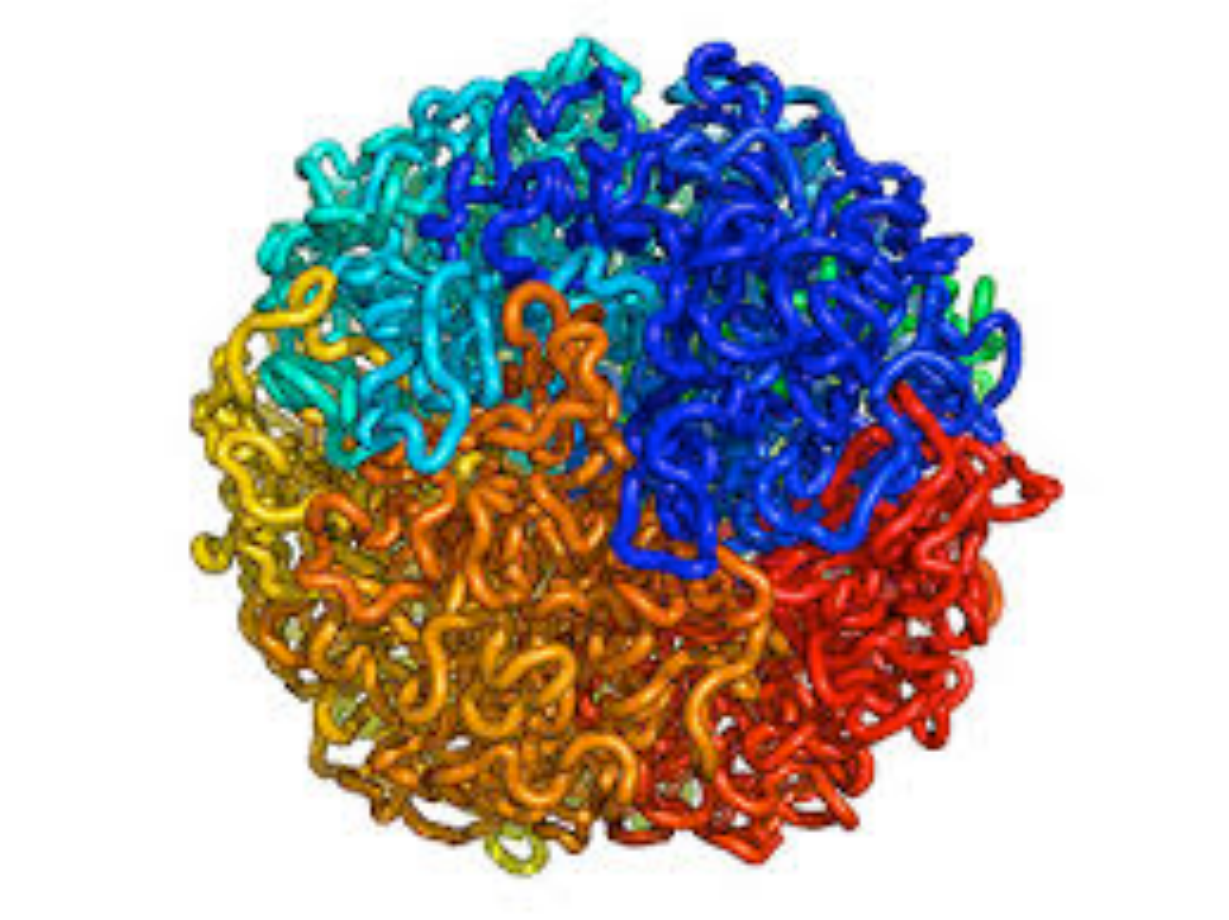
Summary

- Ancestry, i.e. genetic background, matters.
- Collectively, we need to improve the accuracy of “whole” genomes, and also enable the sharing of genotype and phenotype data broadly, among researchers, the research participants and others.
- We need to sequence accurate whole genomes of large pedigrees, and then construct super-family structures.

The End– extra slides to follow



A prisoner at Dartmoor is forced to turn a crank handle repeatedly as a form of punishment, as depicted in an illustration dated 1884.



“There are ~12 billion nucleotides in every cell of the human body, and there are ~25-100 trillion cells in each human body. Given somatic mosaicism, epigenetic changes and environmental differences, no two human beings are the same, particularly as there are only ~7 billion people on the planet”.

Sequencing of 42 genes, captured with Agilent custom capture

The entire coding region, exon-intron boundaries (± 10 bp), and other regions were targeted and captured using Agilent SureSelect custom RNA probes and Integrated DNA Technologies xGen Lockdown custom DNA probes.

Quantified libraries were sequenced on the Illumina MiSeq platform using the 2 x 151 bp configuration to **at least 400x average coverage**. Bioinformatics and data quality control followed the Genome Analysis Toolkit best-practices, with additional algorithms to detect larger insertions, deletions, and duplications.

Conclusion

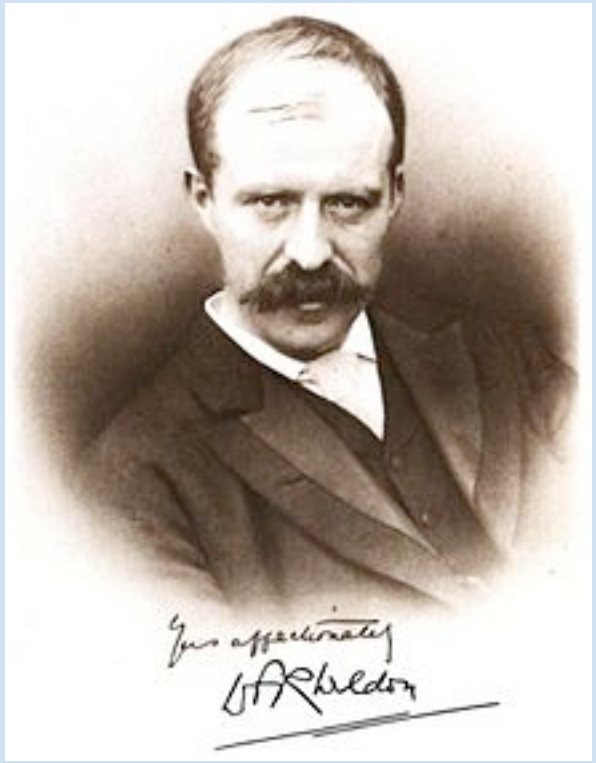
Among women testing negative for *BRCA1/2* mutations, multiple-gene sequencing identified 16 potentially pathogenic mutations in other genes (11.4%; 95% CI, 7.0% to 17.7%), of which 15 (10.6%; 95% CI, 6.5% to 16.9%) prompted consideration of a change in care, enabling early detection of a precancerous colon polyp. Additional studies are required to quantify the penetrance of identified mutations and determine clinical utility. However, these results suggest that multiple-gene sequencing may benefit appropriately selected patients.

J Clin Oncol 32.

Beyond our Kuhnian inheritance

A recent lecture by Prof Greg Radick questions our scientific inheritance, through textbook histories of genetics and Thomas Kuhn's legacy

http://www.guardian.co.uk/science/the-h-word/2012/aug/28/thomas-kuhn



Walter Frank Raphael Weldon

Vs.

William Bateson

Forthcoming by Greg Radick. Scholarly edition of W. F. R. Weldon's Theory of Inheritance (1904-1905), coedited with Annie Jamieson.

Plate I.

Weldon, W. F. R. 1902. Mendel's laws of alternative inheritance in peas. *Biometrika*, 1:228-254.

16p11.2 duplication

Clinical photographs. (e) Proband 5 who has a maternally inherited duplication. (f) Proband 5 (note smooth philtrum) and her healthy duplication positive sister. (g) Duplication positive mother of proband 5, who also has a smooth philtrum. (h) Proband 6 (inherited duplication and oliogohydramnios sequence). Note her frontal bossing, receding hairline, hypoplastic supraorbital ridges and smooth philtrum. (i) Proband 6's right hand showing fifth finger clinodactyly.

Fernandez B A et al. J Med Genet 2010;47:195-203

James Henry Pullen, the idiot savant *who designed the prize winning exhibit for the Paris exhibition in 1867, dressed in the admiral's uniform which he accepted in return for not pursuing his plan to marry. He also designed a realistic model of the* Great Eastern, *a famous transatlantic vessel built by Brunel.*

Master Craftsman
Most famously designed The Great Eastern, a 10 foot long model ship with incredible detail.

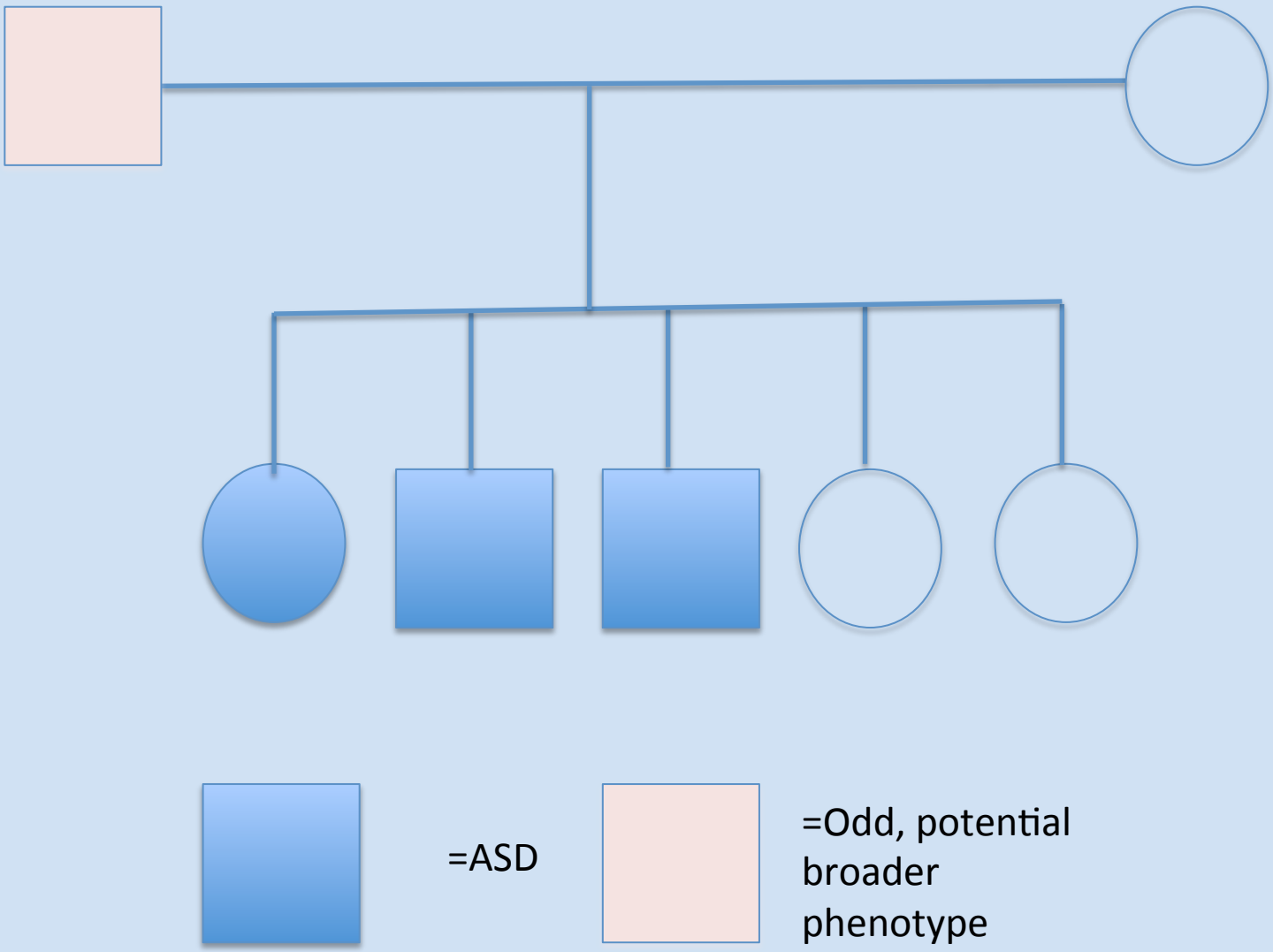
Deaf and nearly mute – Nonverbal,
Obsessed with one topic of building things.
Thought to be mentally retarded.
Usually quiet and reserved, but sometimes was intolerant of advice, suspicious of strangers, and ill-tempered and violent.

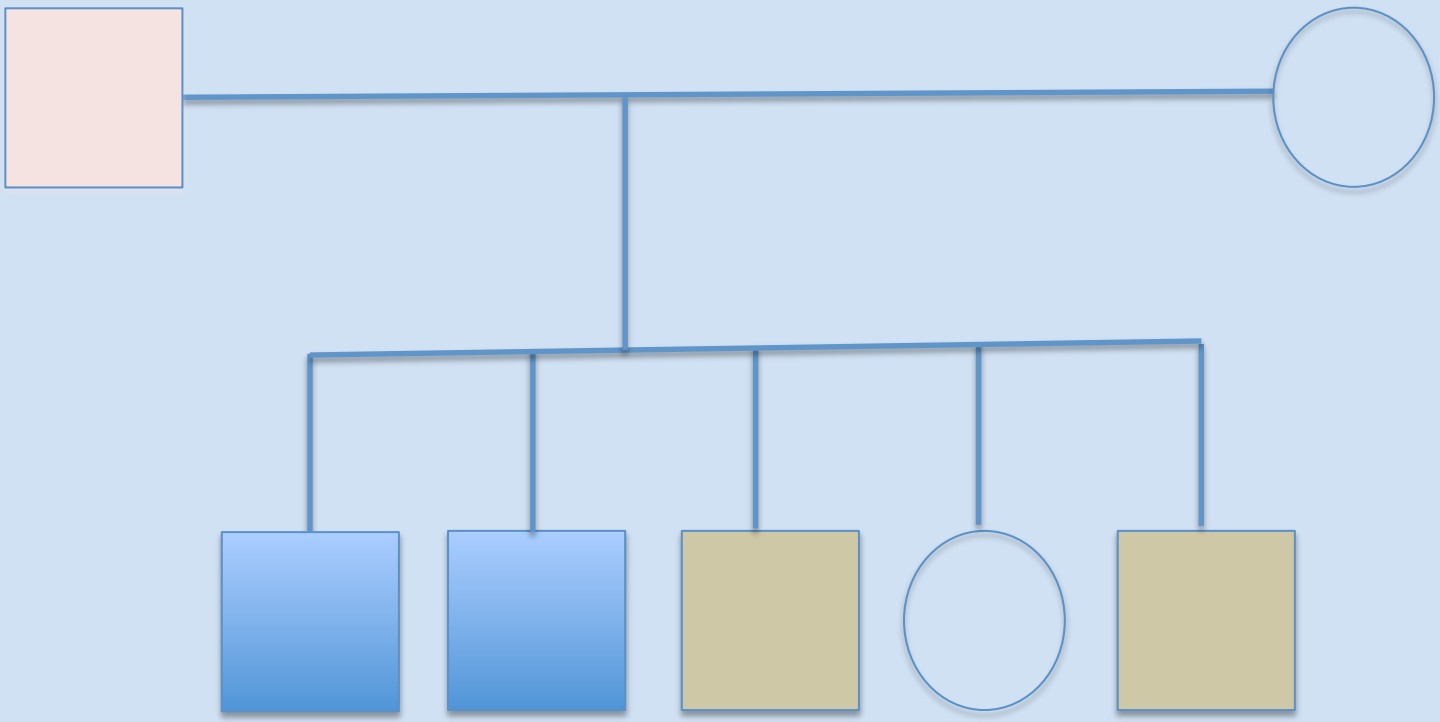
"The clinical and pathological evidence of a pervasive developmental disorder points to a retrospective diagnosis of autism."

Ir J Psych Med 2005; 22(4): 151-155

Sibling Defense Theory

- Defense or modifier Genes? – mutations that somehow protect against or modify the effects of a primary mutation.
- Or, can female gender also somehow be protective with certain mutations?
- Henry Pullen was one of 13 children, but only 3 lived to adult life. His brother, William, was also institutionalized and had exceptional artistic skills. Their parents were first cousins.


=ASD
=Odd, potential broader phenotype

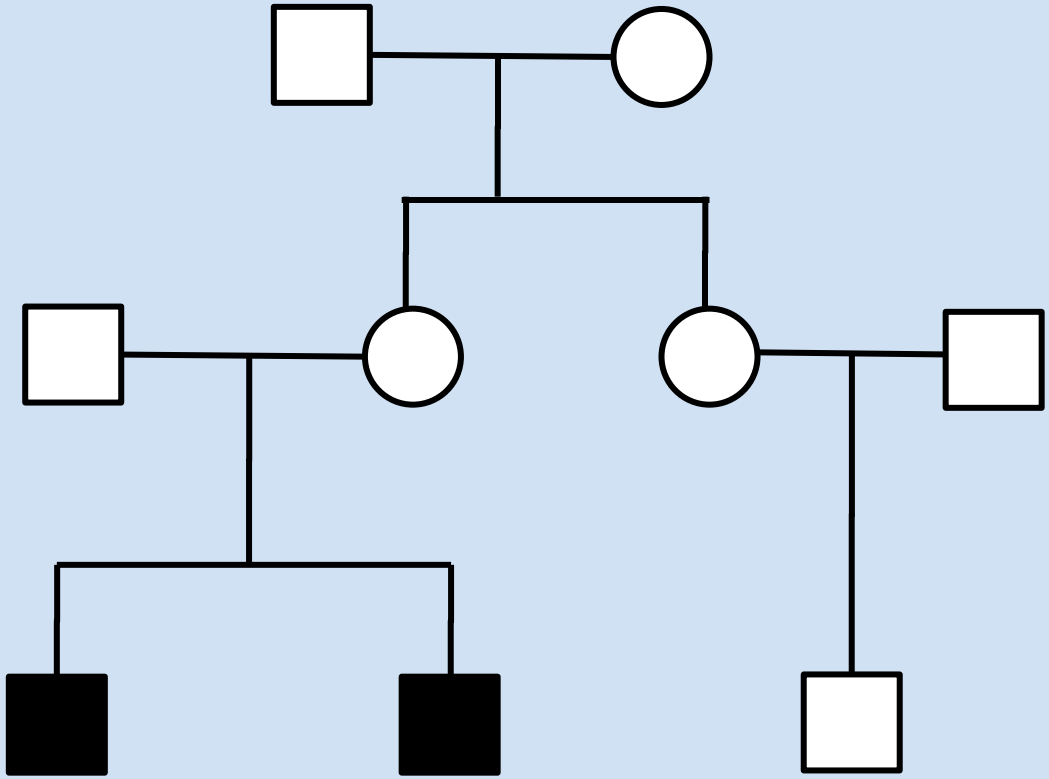


=ASD

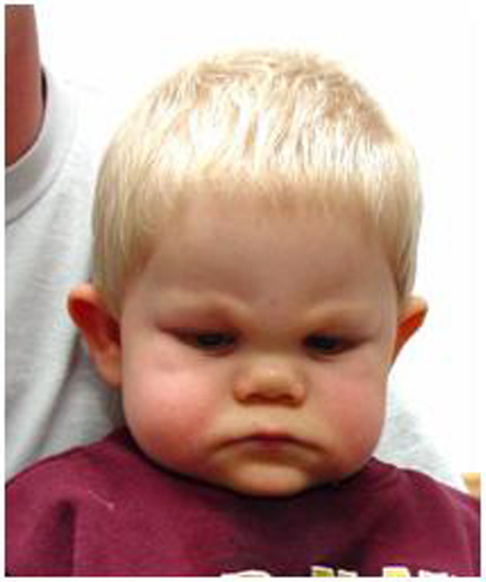
=ADHD

=Odd, potential broader phenotype

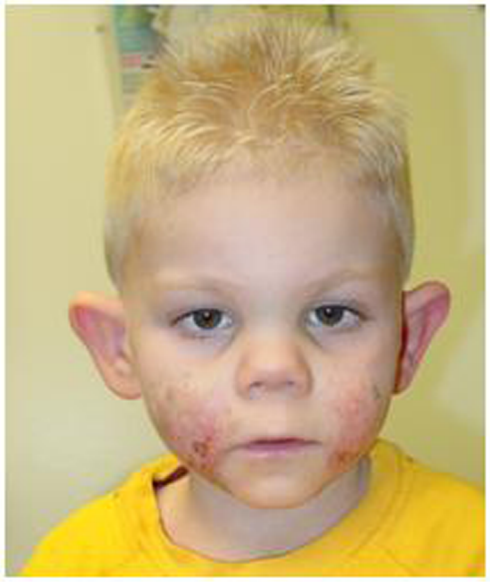
New Syndrome with Mental Retardation, “Autism”, “ADHD”



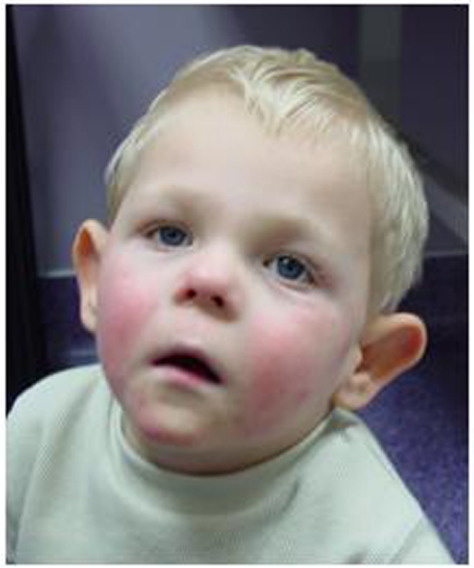
Likely X-linked or Autosomal Recessive, with X-linked being supported by extreme X-skewing in the mother



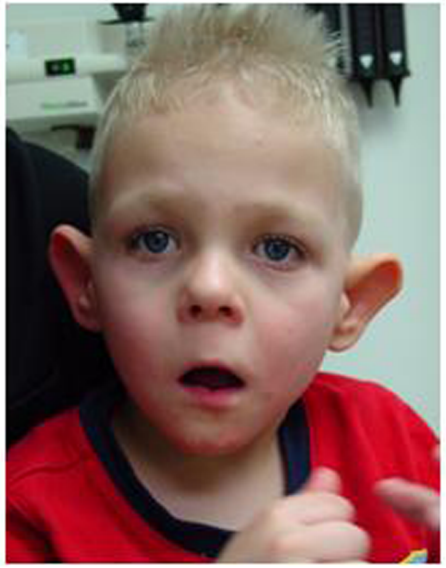
1.5 years old



3.5 years old



3 years old



5 years old

Dysmorphic
Mental Retardation
“autism”
“ADHD”
Hearing difficulties

Workup Ongoing for past 10 years

- Numerous genetic tests negative, including negative for Fragile X and MANY candidate genes.
- Found one missense mutation in a known mental retardation gene, but the mutation is a very conservative nonsynonymous Asp to Glu. Is it relevant or not? What about the whole rest of the genome?

Sequenced whole genomes of Mother, Father and Two Boys, using Complete Genomics

- Sequenced “whole” genomes to obtain noncoding and other non-exonic regions.
- No obvious pathogenic CNVs – microarrays normal.
- ~6 million variants total in the 4 people different from Hg19 reference genome.
- No homozygous autosomal recessive mutations found.
- No Nonsense/Frameshift mutations in both boys.
- 2 mutations present in mother and two boys, on X-chromosome, not in father, not in dbSNP135, not in 1000Genomes April 2012 release, and not in NHLBI 6500 Exomes

2 mutations present in mother and two boys, on X-chromosome, not in father, not in dbSNP135, not in 1000Genomes April 2012 release, and not in NHLBI 6500 Exomes

- Nonsyn SNV ZNF41 c.1191C>A p.Asp397Glu

- Nonsyn SNV TAF1 c.4010T>C p.Ile1337Thr

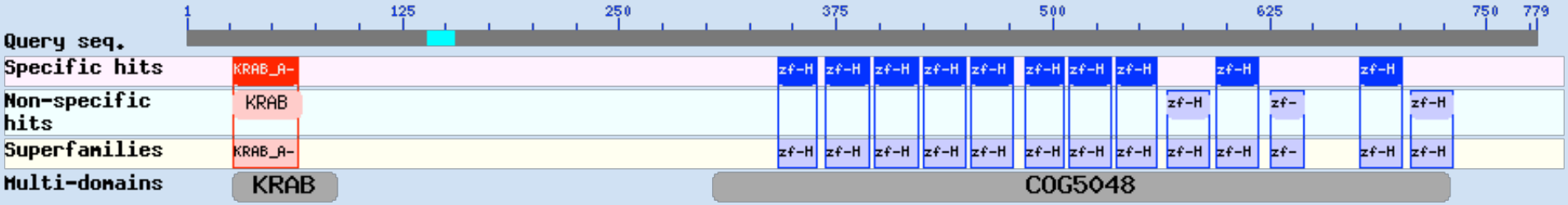
TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 250kDa

Mutations in the *ZNF41* Gene Are Associated with Cognitive Deficits: Identification of a New Candidate for X-Linked Mental Retardation

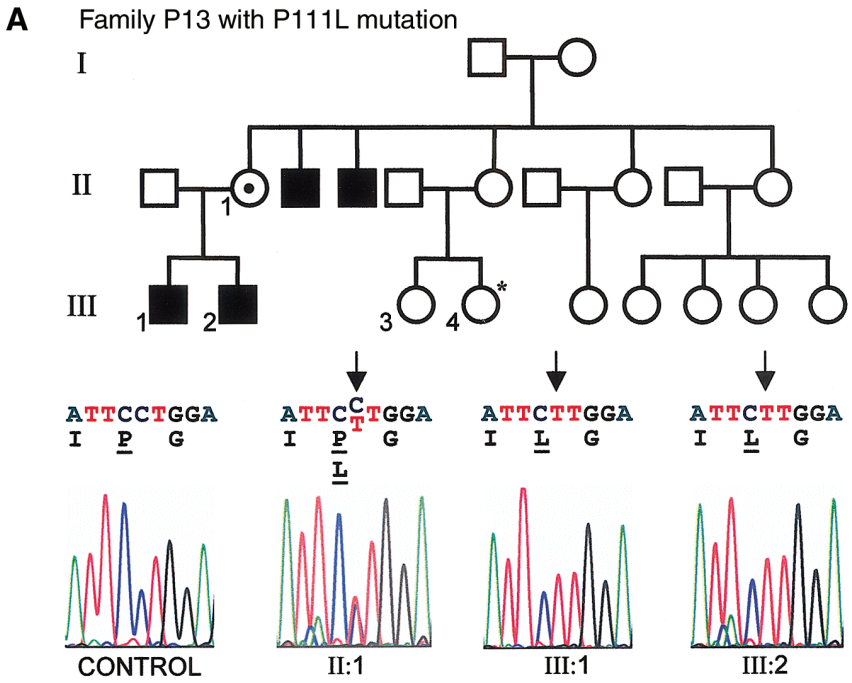
Sarah A. Shoichet,[1] Kirsten Hoffmann,[1] Corinna Menzel,[1] Udo Trautmann,[2] Bettina Moser,[1] Maria Hoeltzenbein,[1] Bernard Echenne,[3] Michael Partington,[4] Hans van Bokhoven,[5] Claude Moraine,[6] Jean-Pierre Fryns,[7] Jamel Chelly,[8] Hans-Dieter Rott,[2] Hans-Hilger Ropers,[1] and Vera M. Kalscheuer[1]

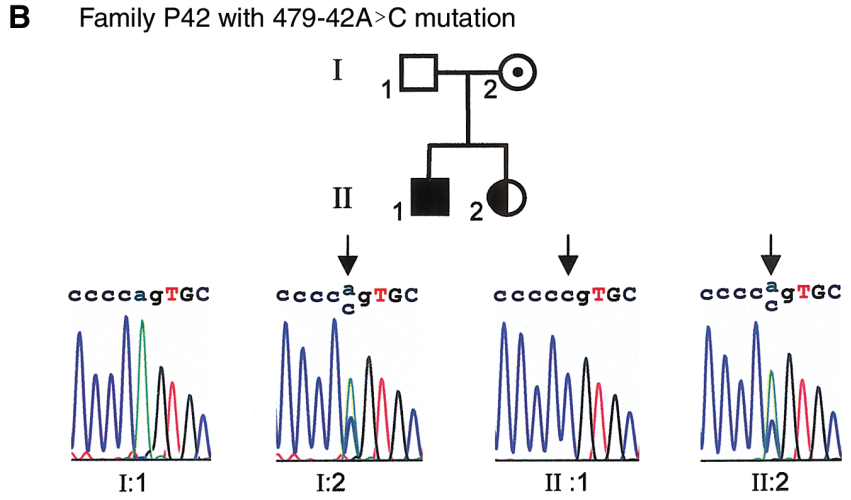
[1]Max-Planck-Institute for Molecular Genetics, Berlin; [2]Institute of Human Genetics, University of Erlangen-Nuremberg, Erlangen-Nuremberg; [3]Centre Hospitalier Universitaire de Montpellier, Hôpital Saint-Eloi, Montpellier, France, [4]Hunter Genetics and University of Newcastle, Waratah, Australia; [5]Department of Human Genetics, University Medical Centre, Nijmegen, The Netherlands; [6]Services de Génétique–INSERM U316, CHU Bretonneau, Tours, France; [7]Center for Human Genetics, Clinical Genetics Unit, Leuven, Belgium; and [8]Institut Cochin de Génétique Moleculaire, Centre National de la Recherche Scientifique/INSERM, CHU Cochin, Paris

Am. J. Hum. Genet. 73:1341–1354, 2003



- KRAB (Kruppel-associated box) domain -A box.
- The KRAB domain is a transcription repression module, found in a subgroup of the zinc finger proteins (ZFPs) of the C2H2 family, KRAB-ZFPs. KRAB-ZFPs comprise the largest group of transcriptional regulators in mammals, and are only found in tetrapods.
- The KRAB domain is a protein-protein interaction module which represses transcription through recruiting corepressors. The KAP1/ KRAB-AFP complex in turn recruits the heterochromatin protein 1 (HP1) family, and other chromatin modulating proteins, leading to transcriptional repression through heterochromatin formation.


A
Family P13 with P111L mutation
I
II
III
1
2
3
4
*
ATTCCTGGA
I P G
ATTCCTTGGA
I P L G
ATTCTTGGA
I L G
ATTCTTGGA
I L G
CONTROL
II:1
III:1
III:2


B
Family P42 with 479-42A>C mutation
I
II
1
2
ccccagTGC
ccccagTGC
ccccgTGC
ccccagTGC
I:1
I:2
II :1
II:2

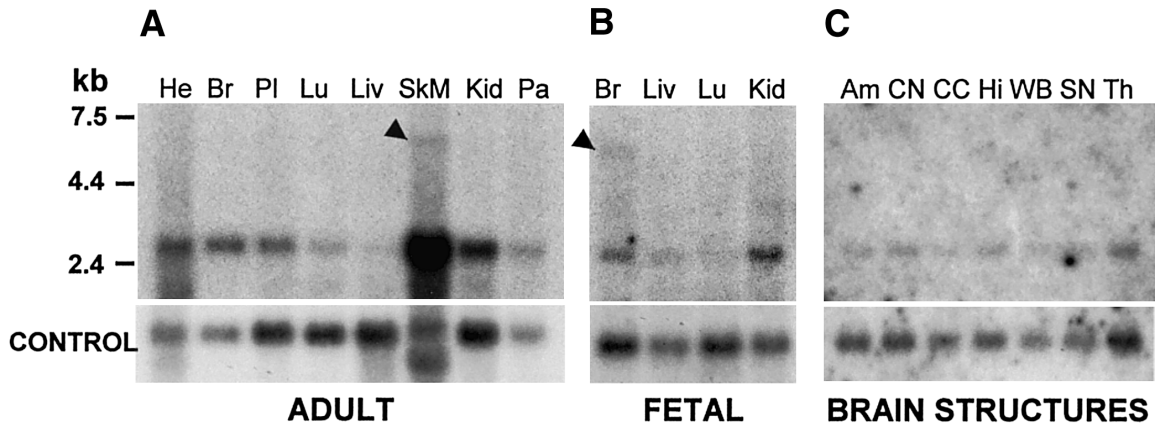


Figure 6 Northern blot hybridization of *ZNF41*, by use of a probe corresponding to nucleotides 621–1099 of *ZNF41* transcript variant 1. *A*, Adult tissues (left to right): heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas. *B*, Fetal tissues (left to right): brain, lung, liver, and kidney. *C*, Adult brain structures (left to right): amygdala, caudate nucleus, corpus callosum, hippocampus, whole brain, substantia nigra, and thalamus. Black arrowheads highlight the presence of a novel 6-kb transcript. *Actin* (*A* and *C*) or *GAPDH* (*B*) served as controls for RNA loading.

Proving Causality

- Will need to find a second, unrelated family with same exact mutation and similar phenotype.
- Can also perform in vitro/in vivo studies and structural modeling, and make knock-in mice and/or test in zebrafish, etc… for biological function.

Genotype First, Phenotype Second AND Longitudinally

Human Molecular Genetics, 2010, Vol. 19, Review Issue 2 ***R176–R187***
doi:10.1093/hmg/ddq366
Advance Access published on August 31, 2010

Phenotypic variability and genetic susceptibility to genomic disorders

Santhosh Girirajan and Evan E. Eichler*

Department of Genome Sciences, Howard Hughes Medical Institute, University of Washington School of Medicine, PO Box 355065, Foege S413C, 3720 15th Avenue NE, Seattle, WA 98195, USA

Genome-Wide Association Study of Multiplex Schizophrenia Pedigrees

Am J Psychiatry Levinson et al.; AiA:1–11

"Rare CNVs were observed in regions with strong previously documented association with schizophrenia, but with variable patterns of segregation. This should serve as a reminder that we still know relatively little about the distribution of these CNVs in the entire population (e.g., in individuals with no or only mild cognitive problems) or about the reasons for the emergence of schizophrenia in only a minority of carriers, so great caution is required in genetic counseling and prediagnosis."

Clinical Management and Genetics

Genet Med. 2011 Sep;13(9):770-6.

Chromosomal microarray testing influences medical management.

Coulter ME, Miller DT, Harris DJ, Hawley P, Picker J, Roberts AE, Sobeih MM, Irons M.

PURPOSE:

Chromosomal microarray (CMA) testing provides the highest diagnostic yield for clinical testing of patients with developmental delay (DD), intellectual disability (ID), multiple congenital anomalies (MCA), and autism spectrum disorders (ASD). Despite improved diagnostic yield and studies to support cost-effectiveness, concerns regarding the cost and reimbursement for CMA have been raised because it is perceived that CMA results do not influence medical management.

METHODS:

We conducted a retrospective chart review of CMA testing performed during a 12-month period on patients with DD/ID, ASD, and congenital anomalies to determine the proportion of cases where abnormal CMA results impacted recommendations for clinical action.

RESULTS:

Among 1792 patients, 13.1% had clinically relevant results, either abnormal (n = 131; 7.3%) or variants of possible significance (VPS; n = 104; 5.8%). Abnormal variants generated a higher rate of recommendation for clinical action (54%) compared with VPS (34%; Fisher exact test, P = 0.01). CMA results influenced medical care by precipitating medical referrals, diagnostic imaging, or specific laboratory testing.

CONCLUSIONS:

For all test indications, CMA results influenced medical management in a majority of patients with abnormal variants and a substantial proportion of those with VPS. These results support the use of CMA as a clinical diagnostic test that influences medical management for this patient population.