Challenges for Clinical Implementation of Genomic Medicine

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

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

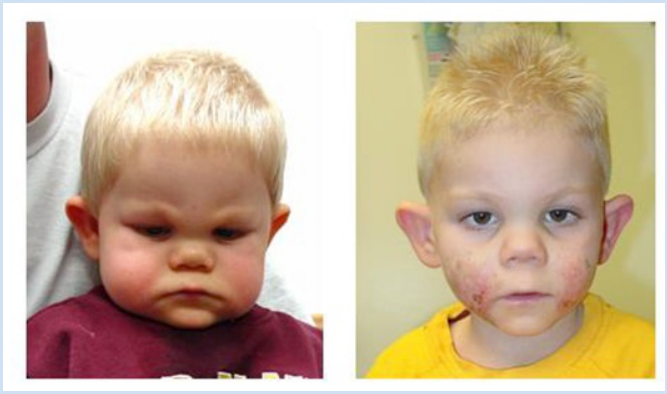
Conflicts of Interest

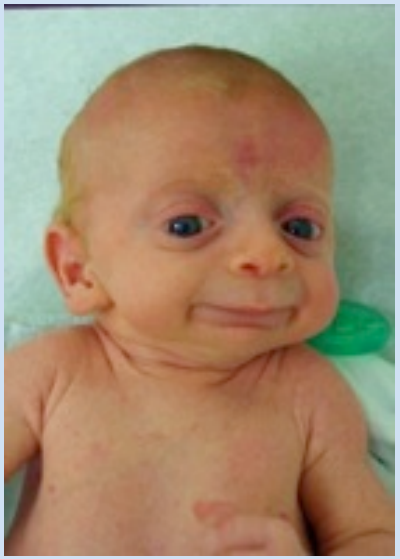
Advisory Boards

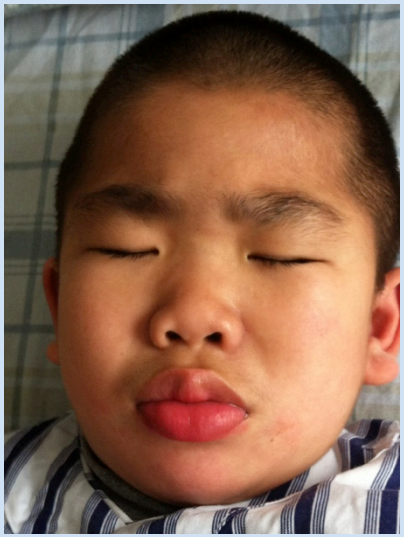
genepeeks

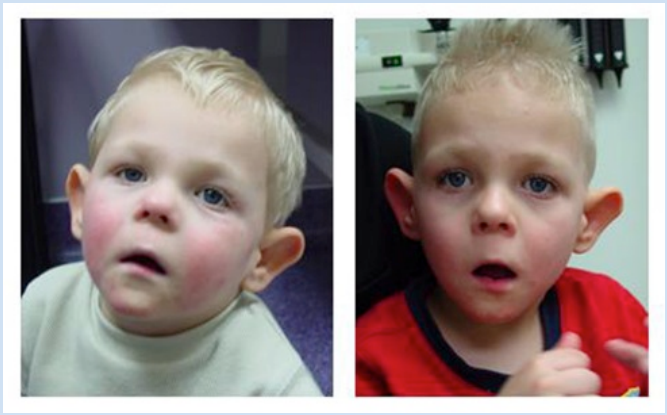
Omicia
Unlocking Individualized Medicine

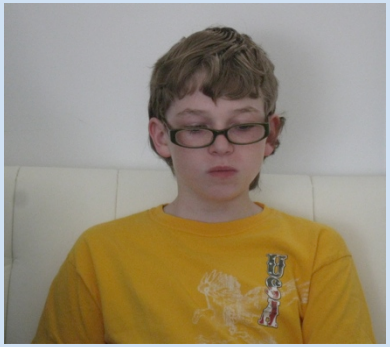
Rare Diseases



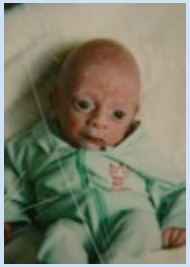




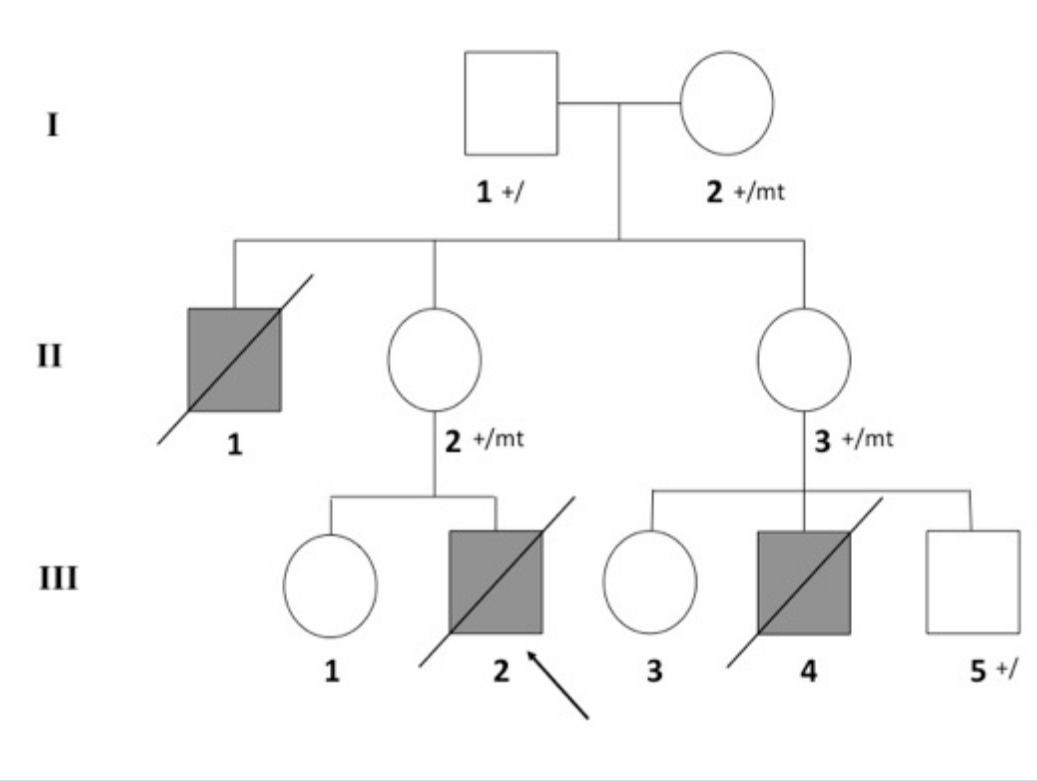


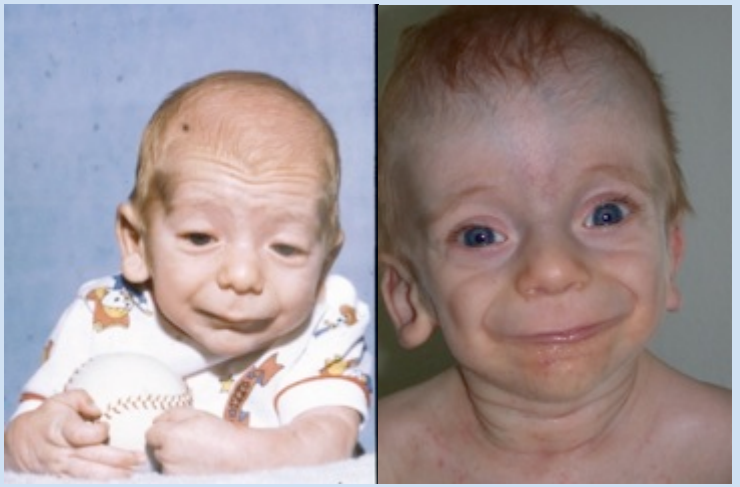


One family in Utah with a very rare disease.



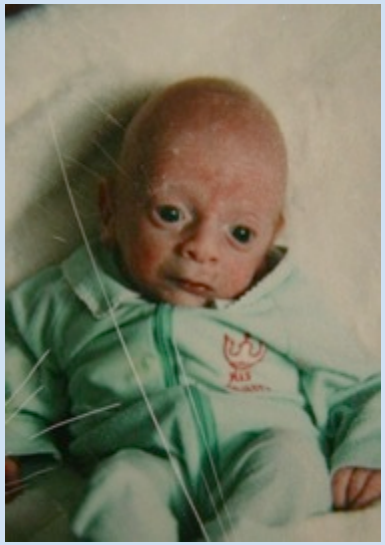
An unrelated second family was also identified, due to sharing the same genotype, i.e. the same mutation.

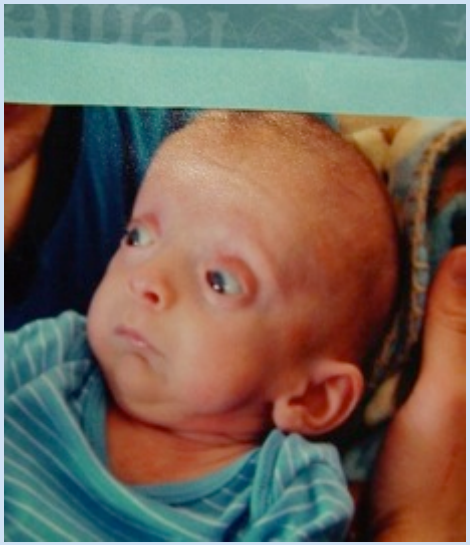
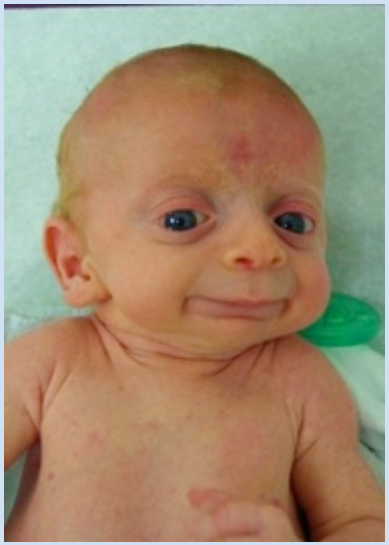
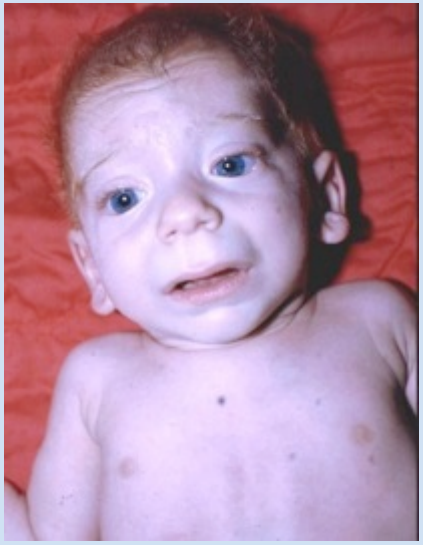
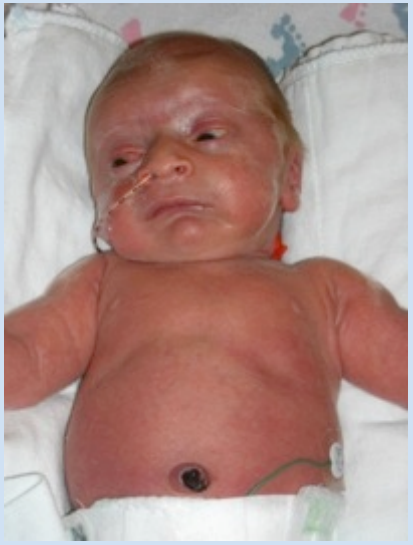
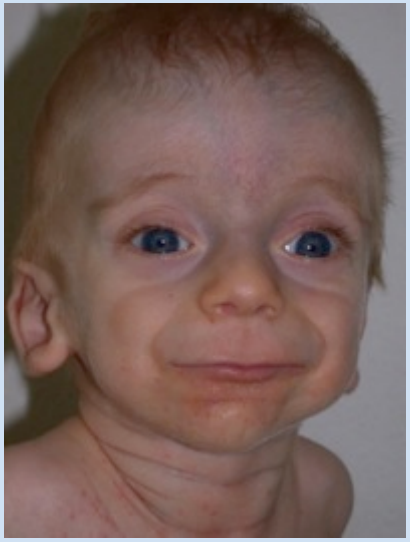




II-1 III-2

Ogden Syndrome, in honor of where the first family lives, in Ogden, Utah

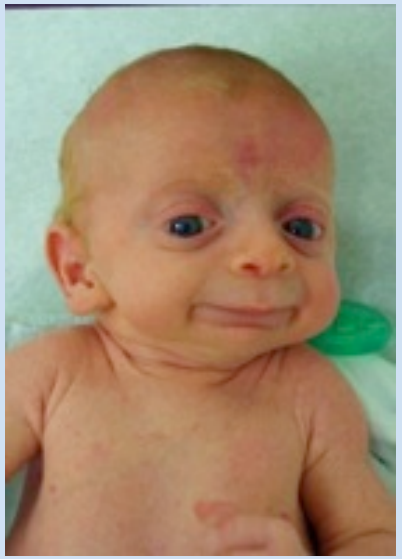


Big Questions though:

Simulated structure of S37P mutant

?



What is the molecular basis of Ogden syndrome?

- Naa10/Naa15 complex
- Naa10 localisation
- Naa10 function

what can we learn from Ogden syndrome?

- characterizing different model systems (fibroblasts, yeast, *C. elegans)*

patient cells

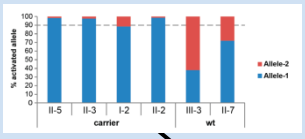
molecular level

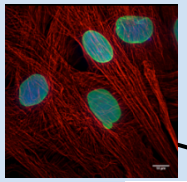
C. elegans

HEK293 cells

Naa10 S37P

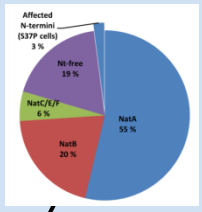
X-chromosome skewing in carriers



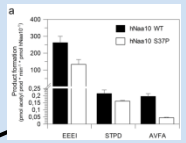


normal nuclear envelope

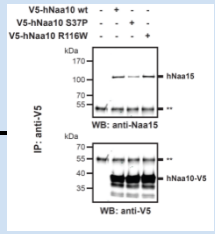
affected *in vivo* substrates

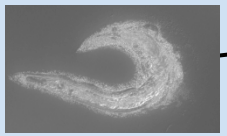


reduced enzymatic activity



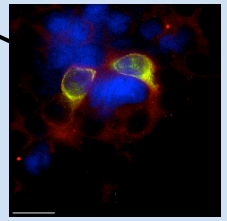
disrupted NatA complex





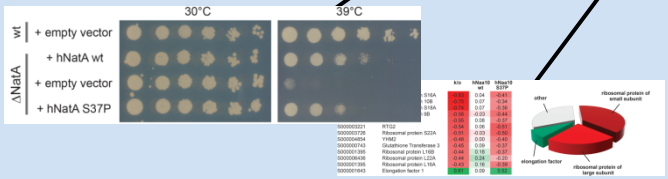
initial experiments

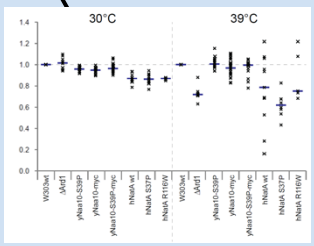
normal localization



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

S39 mutation?
growth defect of “humanized” strains





S. cerevisiae

Challenges for Clinical Implementation of Genomic Medicine

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

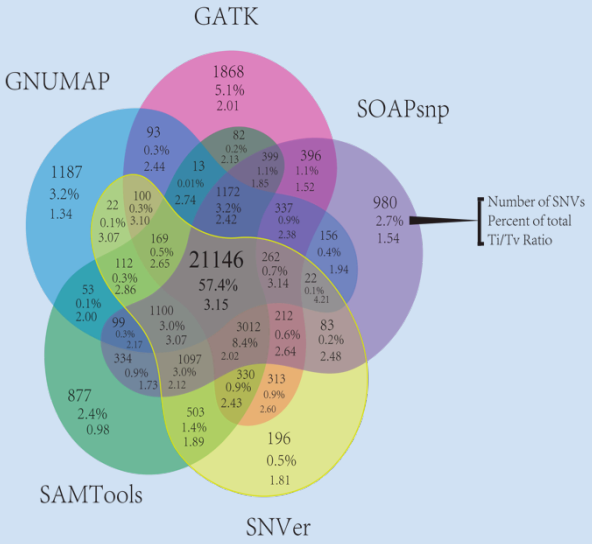
Genome Medicine

RESEARCH Open Access

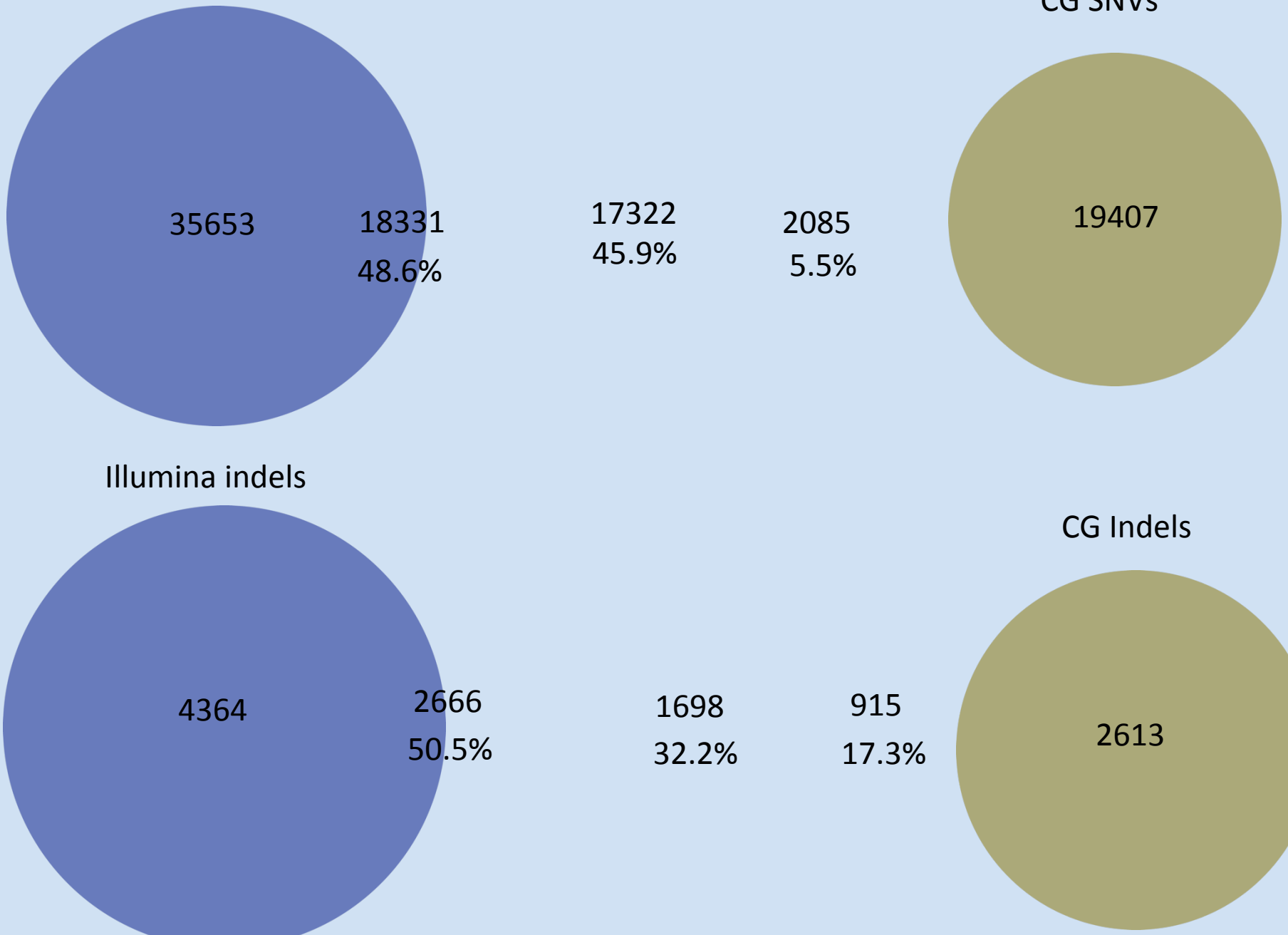
Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing

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?



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

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.

Limits of our current technology & knowledge

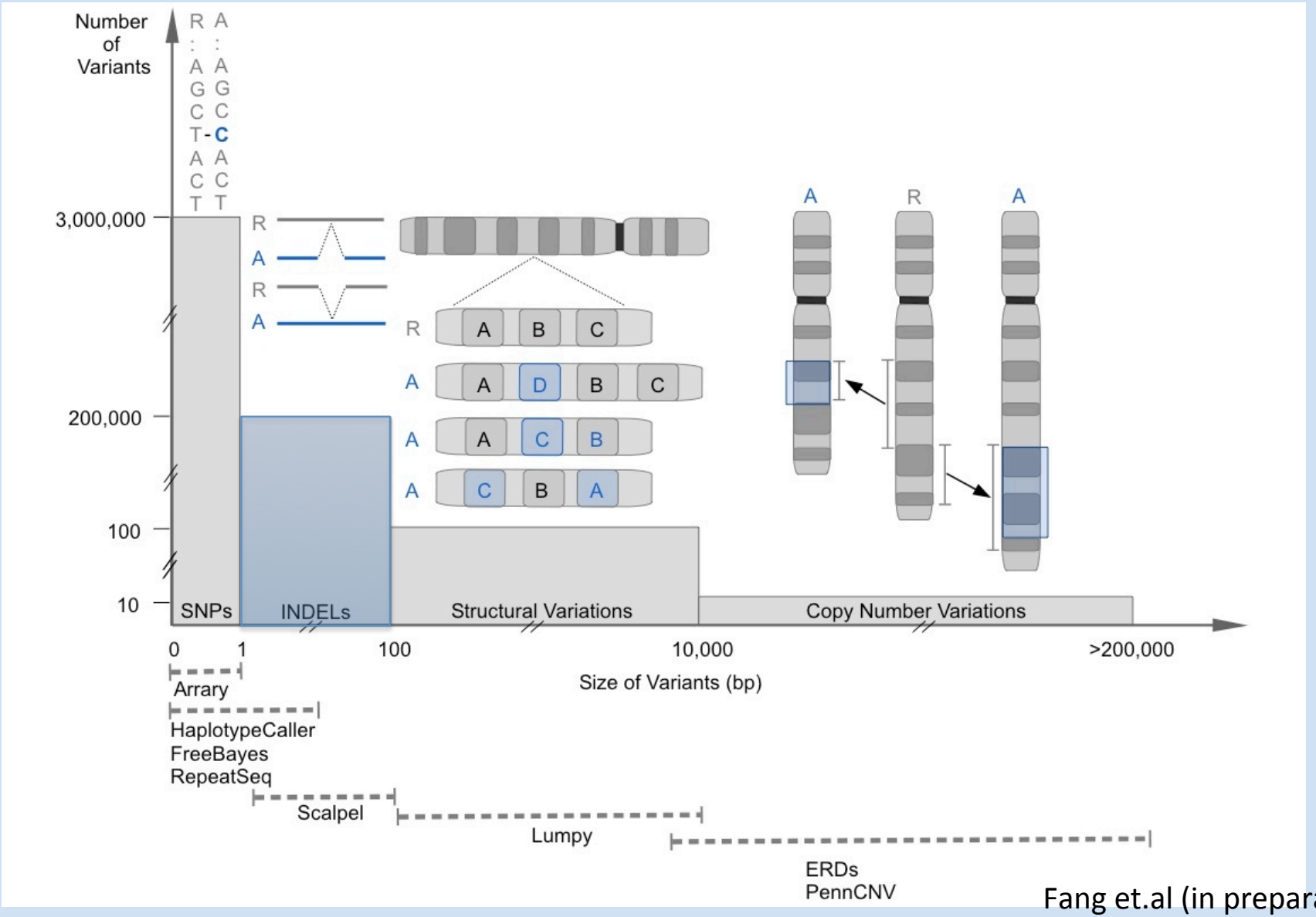
Analytic Validity

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

Clinical Validity

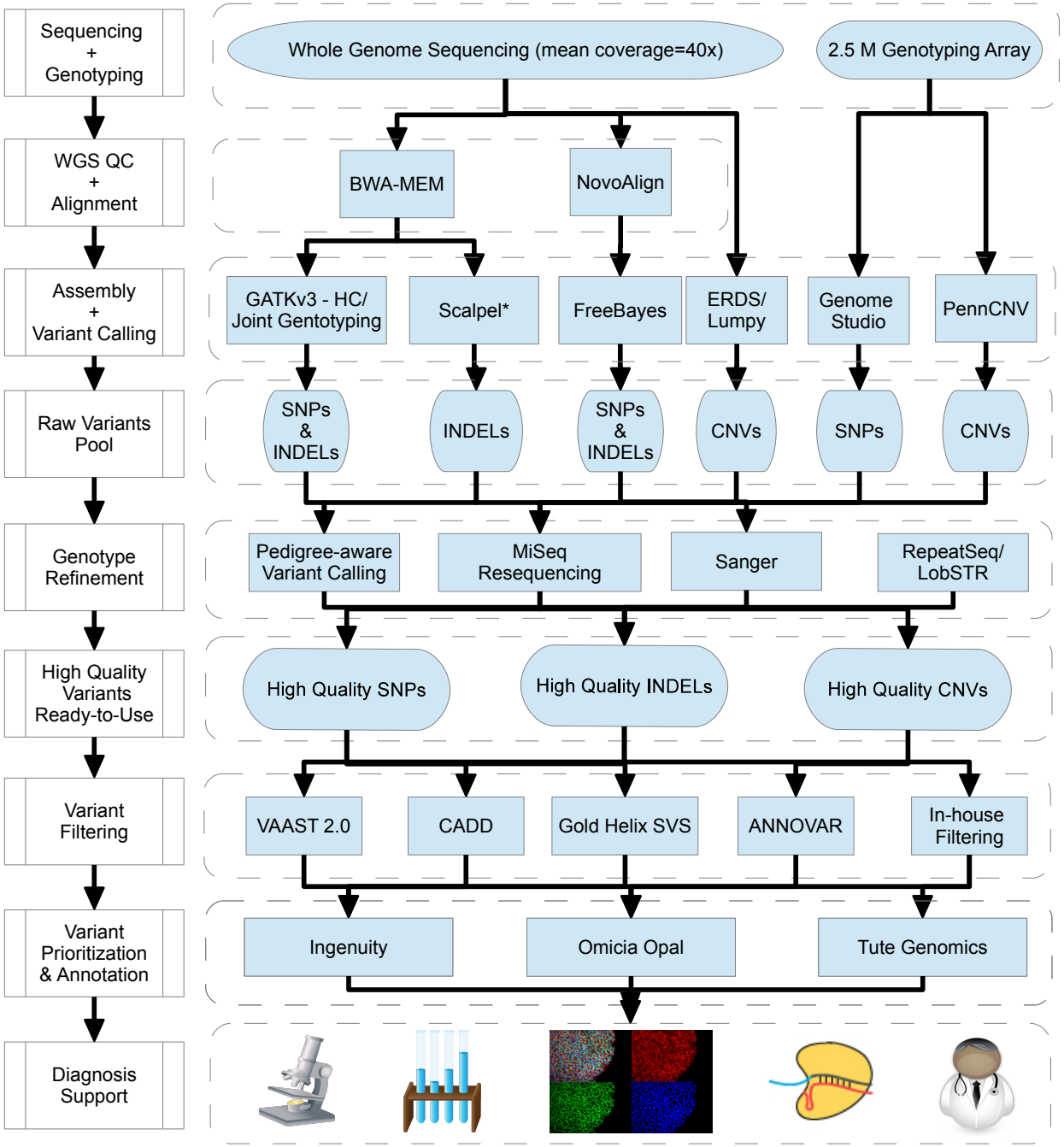
- Genetic architecture of illness

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

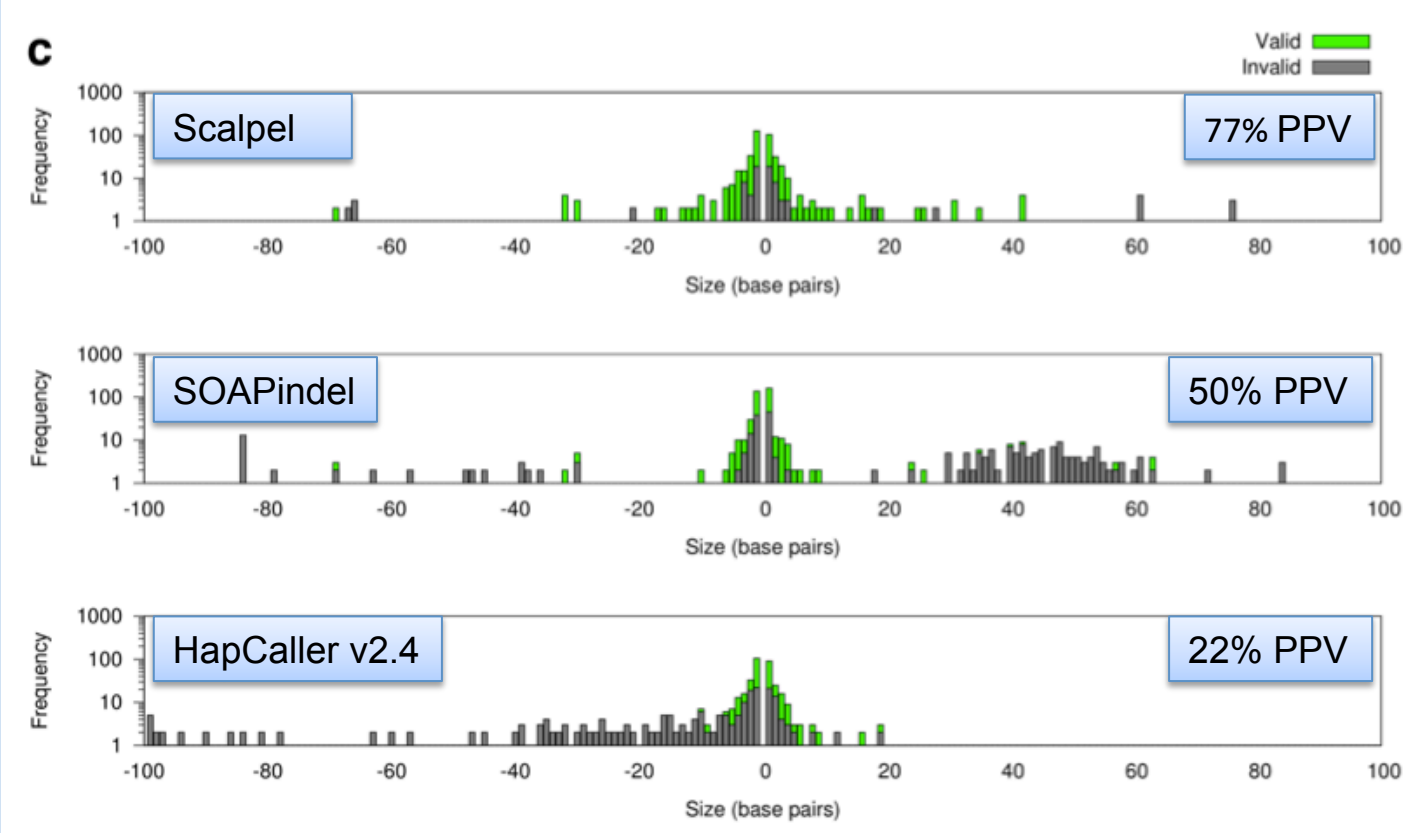
Accurate detection of de novo and transmitted INDELs within exome-capture data using micro-assembly

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



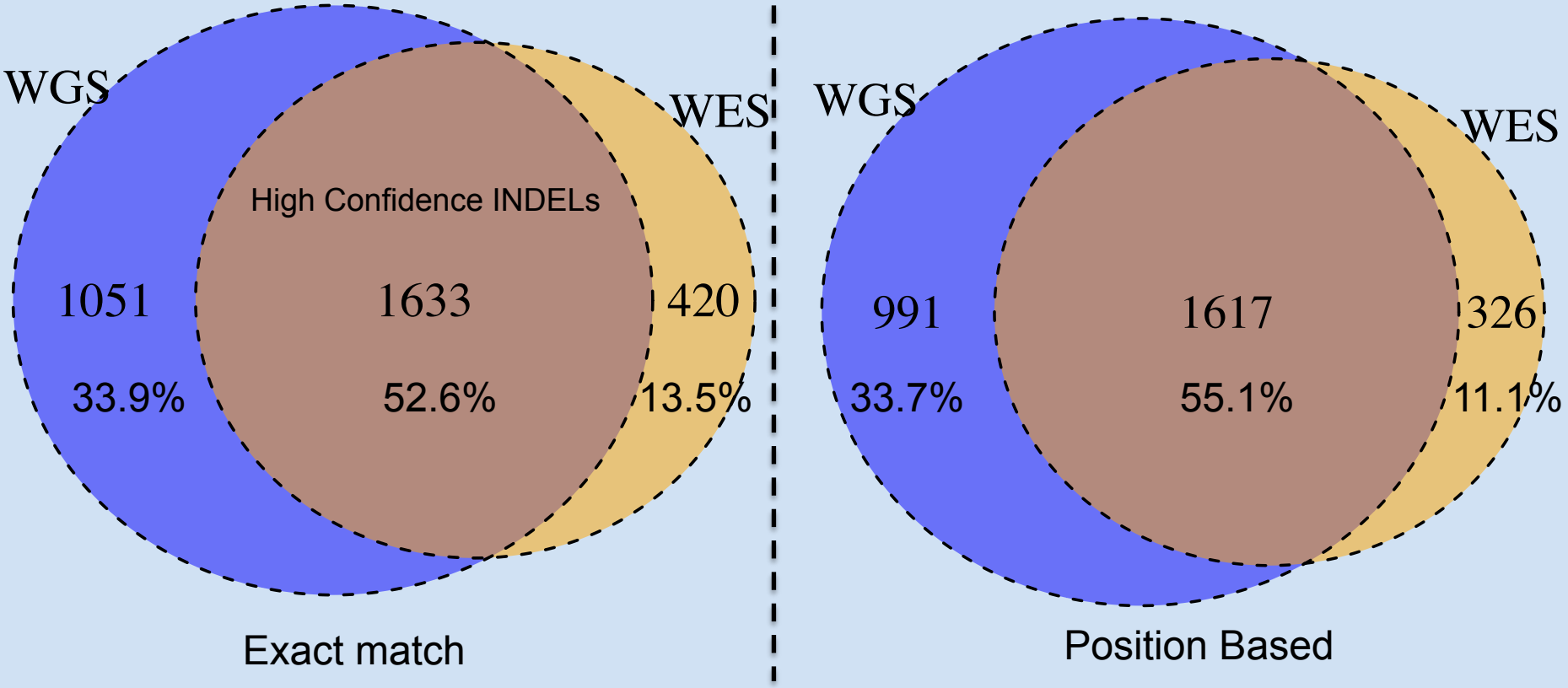
There are recent improvements with GATK v3.0 with 45% PPV, but Scalpel still out-performs this.

Narzisi et.al (Accepted in Nat. Methods)

Extending Scalpel with comparisons of WGS & WES data

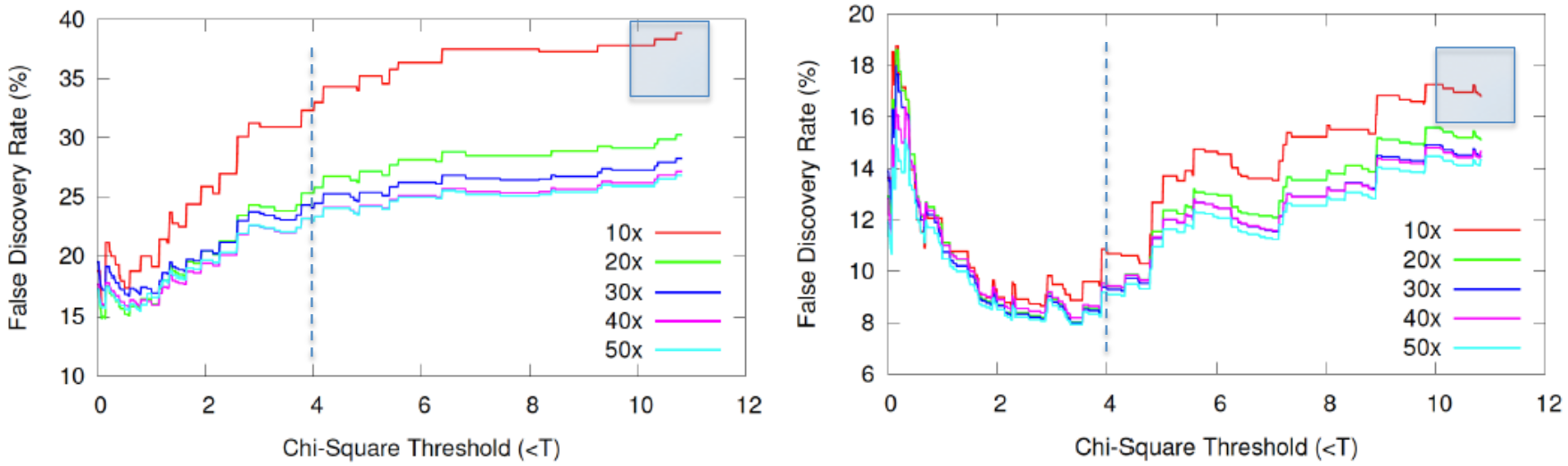
- WGS and WES were performed on 8 samples.
- Illumina HiSeq 2000 platform, paired-end 100 bp reads.
- Exome Capture Kit: NimbleGen SeqCap EZ Exome v2.0 capture reagent, representing 36.0 Mb (approximately 300,000 exons) of the human genome (hg19 build).
- WGS: Mean coverage= ~70x, ~95% > 20x
- WES: Mean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

Validation data showed that low confidence INDELs have higher Chi-Square scores and lower coverage for the alternative allele.



Characterization of False Discovery Rate as a function of the chi-squared score and coverage. The analysis is reported for mutations outside of microsatellites (left plot) and within microsatellites (right plot). Different curves are plotted for subsets of mutations with predefined maximum coverage.

$$X^2 = \frac{(C_o^R - C_e^R)^2}{C_e^R} + \frac{(C_o^A - C_e^A)^2}{C_e^A}$$

Definitions:

Low error INDELs: Coverage(alternative allele) >10 or χ^2 <4

High error INDELs: Coverage(alternative allele) <10 and χ^2 >10.84

Moderate confidence INDELs: Do not fall into the above two categories.

Note: C: k-mer coverage counts, o: observed e: expected, R: Reference A: Alternative

INDELs in the validation: ~650 out of a total of 2300 INDELs

Narzis et.al (Accepted in Nat. Methods)

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

Figures were removed for posting.

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?

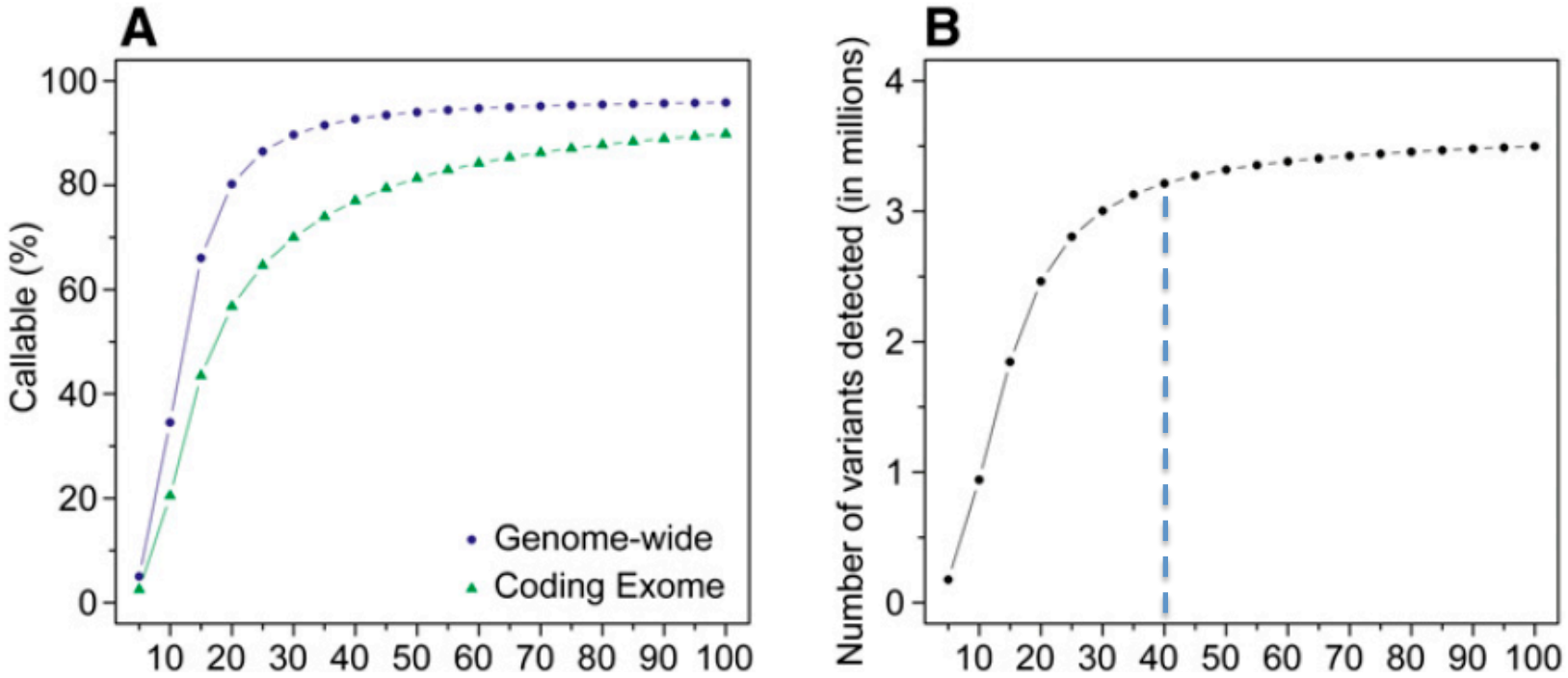


Figure 5. Genotype calling as a function of average mapped depth. The *x*-axes represent the average mapped depth of each data set, and the *y*-axes represent the proportion of the whole genome (dark blue circles) and coding exome (green triangles) that is callable (*A*), the number of SNVs detected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

Figures were removed for posting.

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

Figures were removed for posting.

Limits of our current technology & knowledge

Analytic Validity

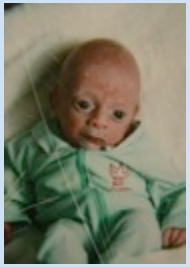
- Sequencing “clinical-grade genomes”

- Bioinformatics analysis

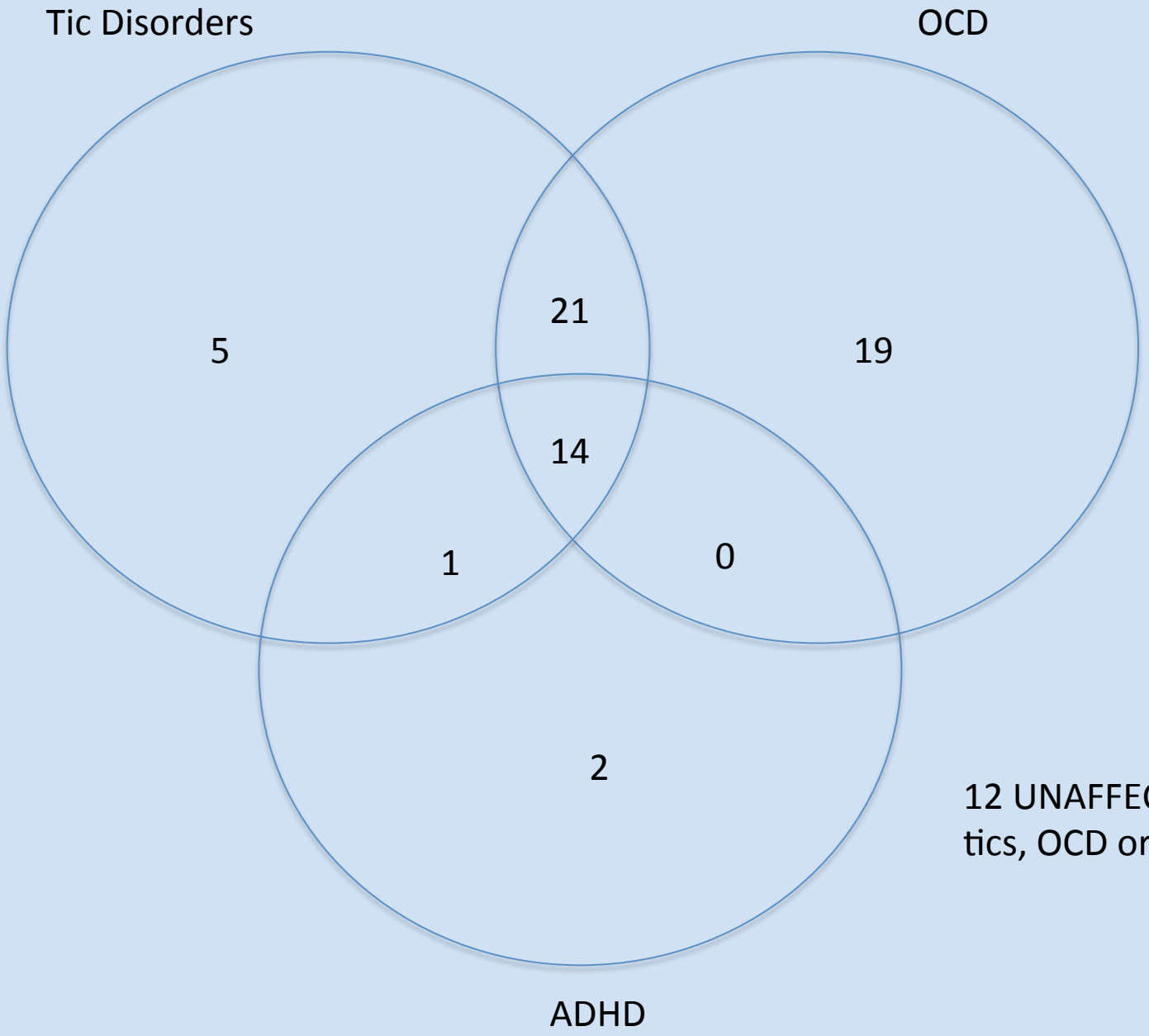
Clinical Validity

- **Genetic architecture of illness**

Ogden Syndrome

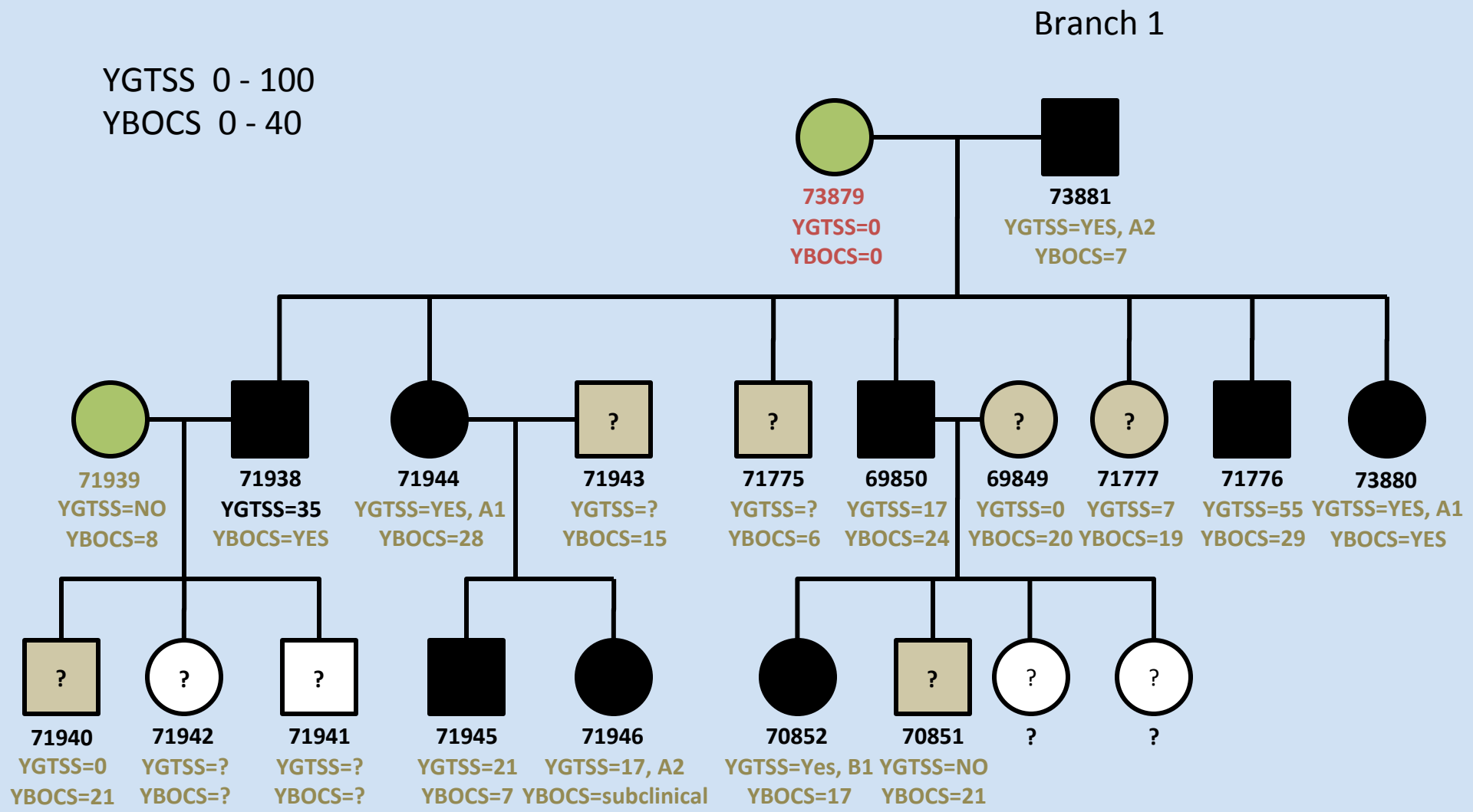


A Large Pedigree in Utah with mental illness



12 UNAFFECTED with tics, OCD or ADHD

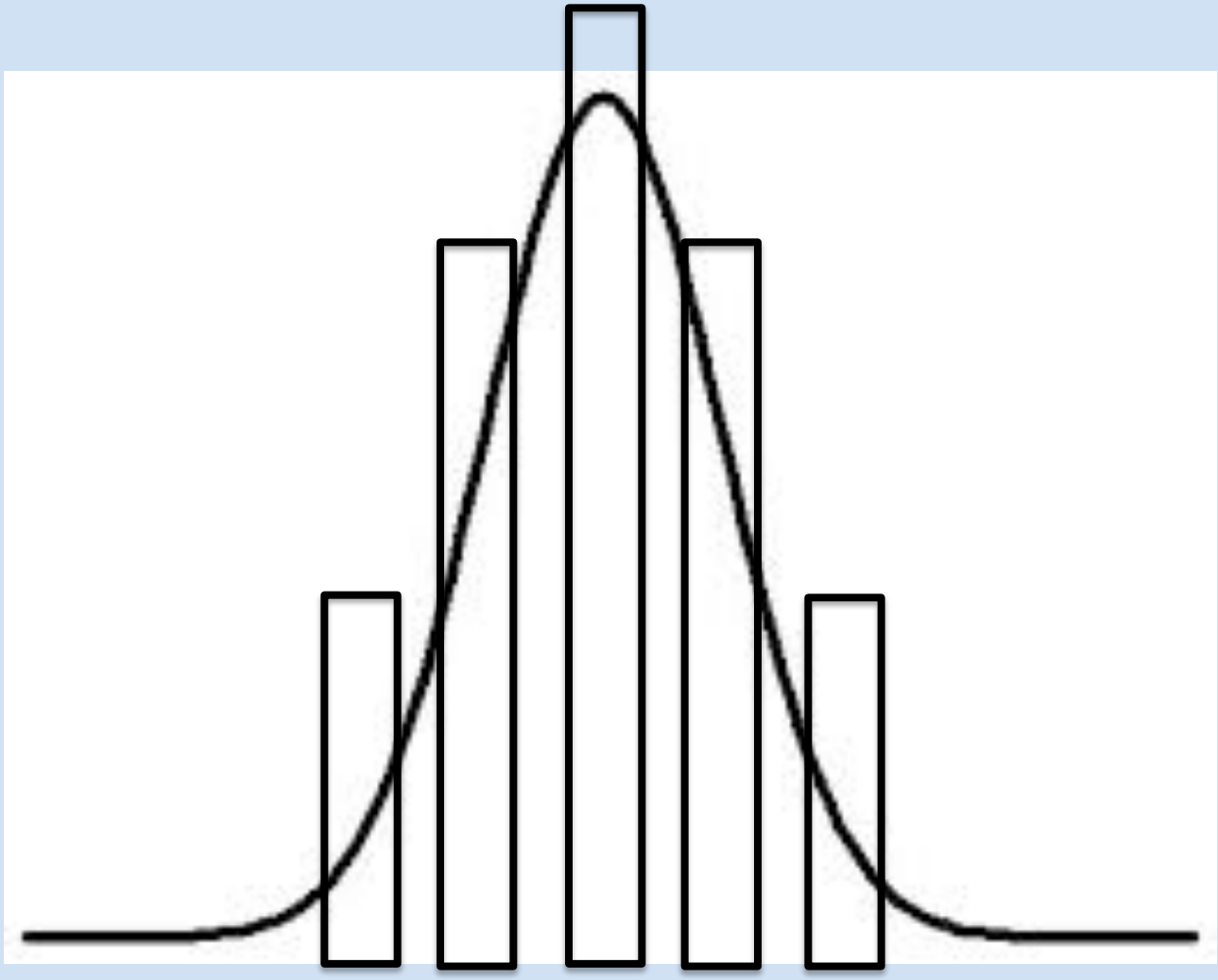
Phenotyping of just one branch in this pedigree

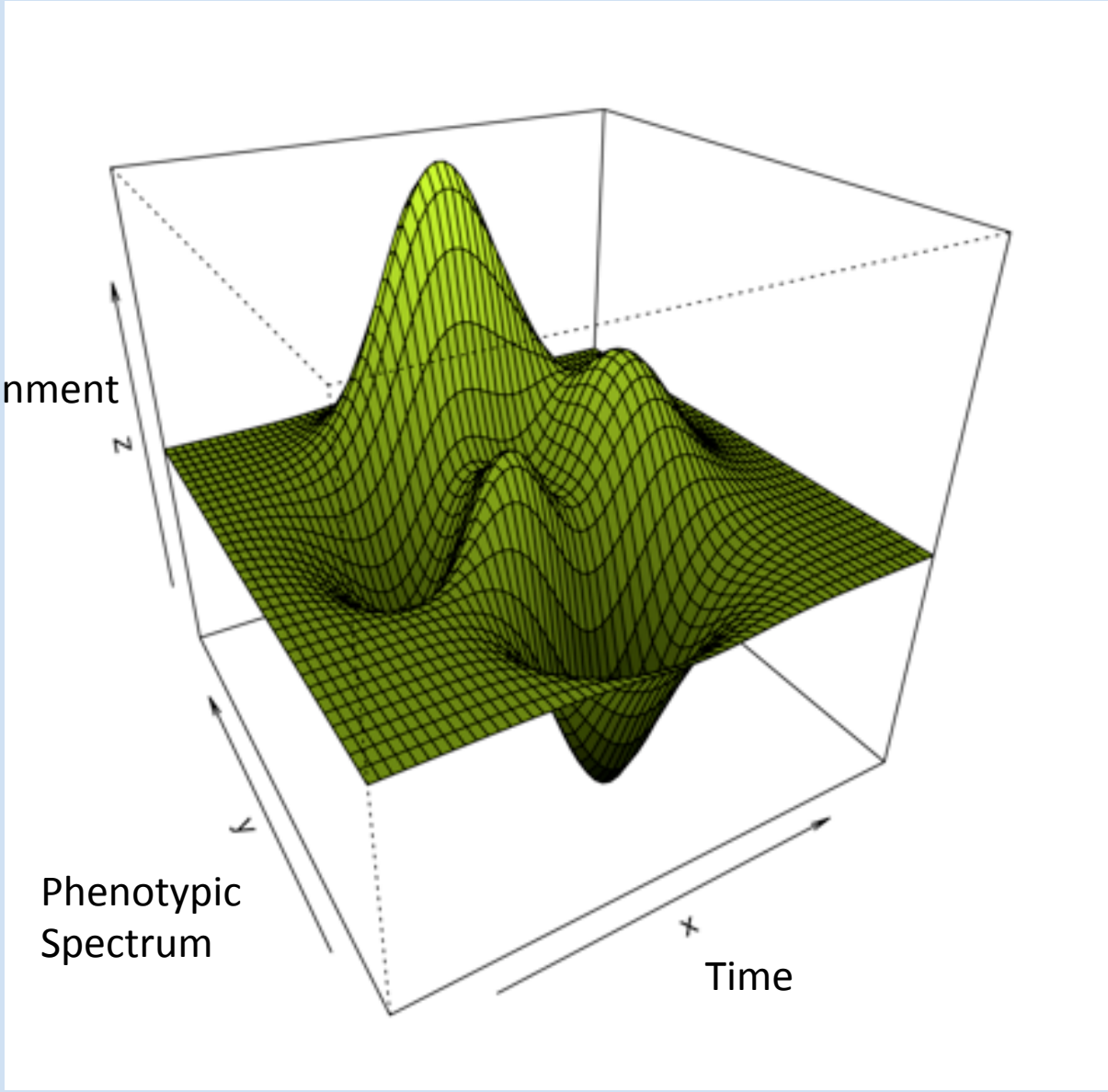


Expression Issues

- 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.

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

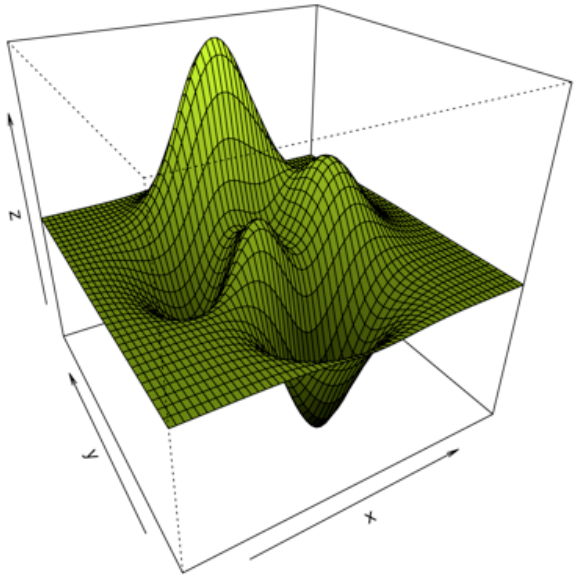
XLID-Causing Mutations and Associated Genes Challenged in Light of Data From Large-Scale Human Exome Sequencing

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 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

PERSPECTIVE

OPEN
doi:10.1038/nature13127

Guidelines for investigating causality of sequence variants in human disease

D. G. MacArthur[1,2], T. A. Manolio[3], D. P. Dimmock[4], H. L. Rehm[5,6], J. Shendure[7], G. R. Abecasis[8], D. R. Adams[9,10], R. B. Altman[11], S. E. Antonarakis[12,13], E. A. Ashley[14], J. C. Barrett[15], L. G. Biesecker[16], D. F. Conrad[17], G. M. Cooper[18], N. J. Cox[19], M. J. Daly[1,2], M. B. Gerstein[20,21], D. B. Goldstein[22], J. N. Hirschhorn[2,23], S. M. Leal[24], L. A. Pennacchio[25,26], J. A. Stamatoyannopoulos[27], S. R. Sunyaev[28,29], D. Valle[30], B. F. Voight[31], W. Winckler[2]† & C. Gunter[18]†

The discovery of rare genetic variants is accelerating, and clear guidelines for distinguishing disease-causing sequence variants from the many potentially functional variants present in any human genome are urgently needed. Without rigorous standards we risk an acceleration of false-positive reports of causality, which would impede the translation of genomic research findings into the clinical diagnostic setting and hinder biological understanding of disease. Here we discuss the key challenges of assessing sequence variants in human disease, integrating both gene-level and variant-level support for causality. We propose guidelines for summarizing confidence in variant pathogenicity and highlight several areas that require further resource development.

Clinical Validity?

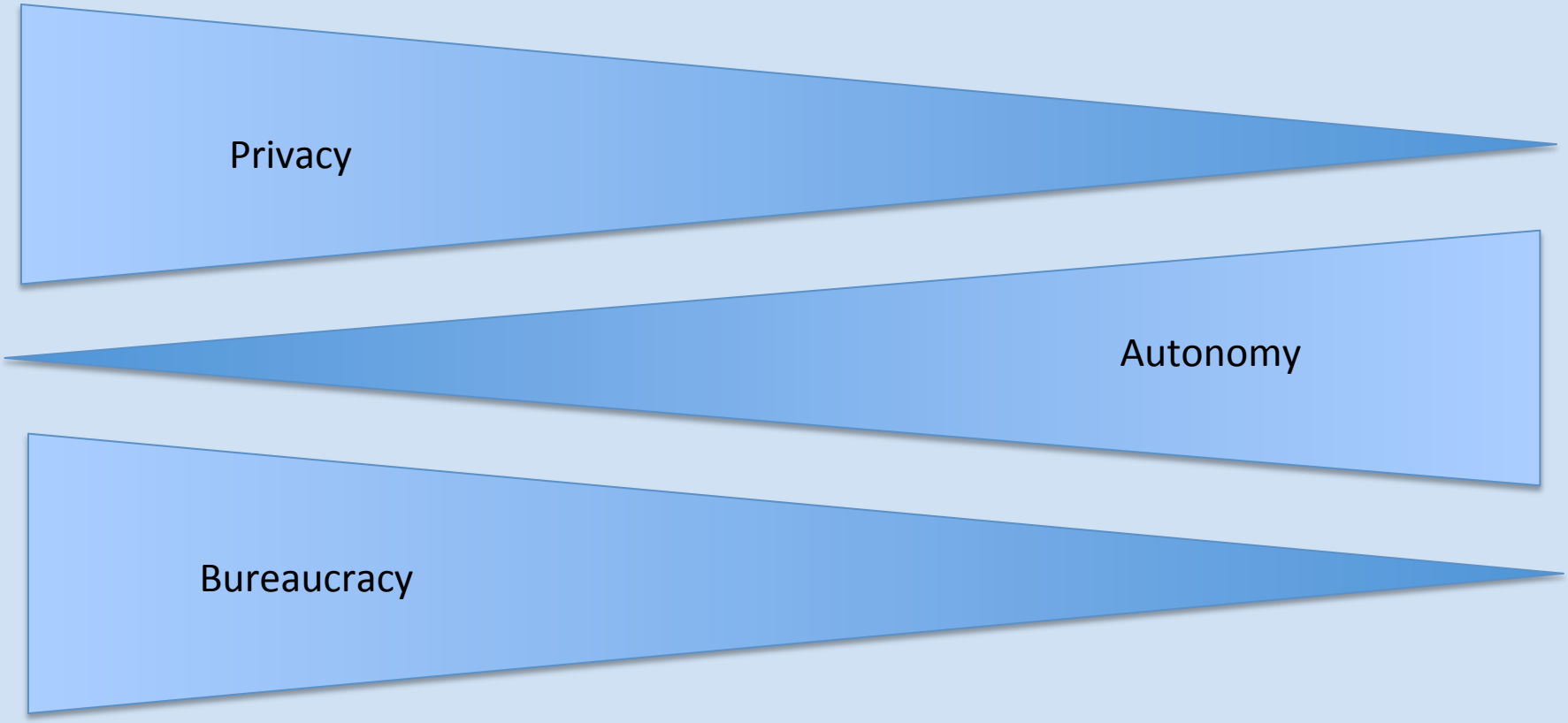
This is SO complex that the only solid way forward is with a “networking of science” model, i.e. online database with genotype and phenotype longitudinally tracked for thousands of volunteer families.

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

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

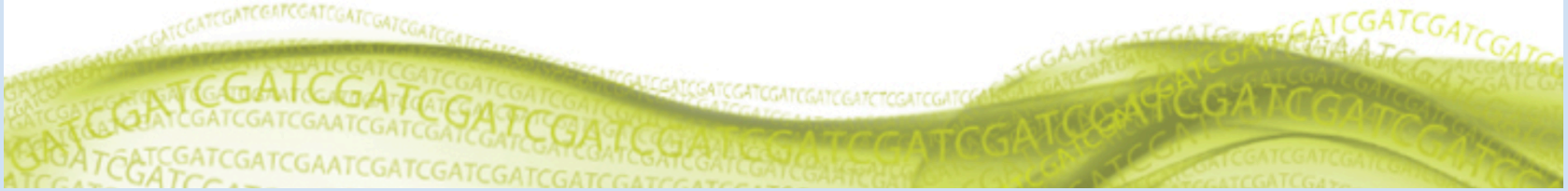
Individual Genome Sequencing Service

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

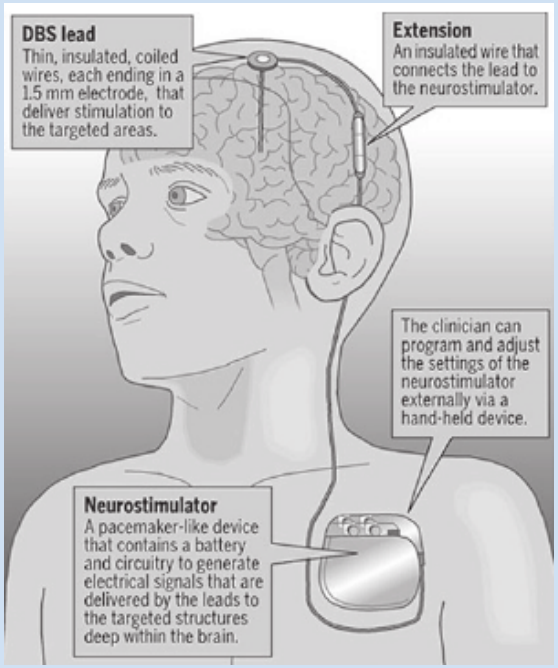
www.everygenome.com
CLIA#: 05D1092911

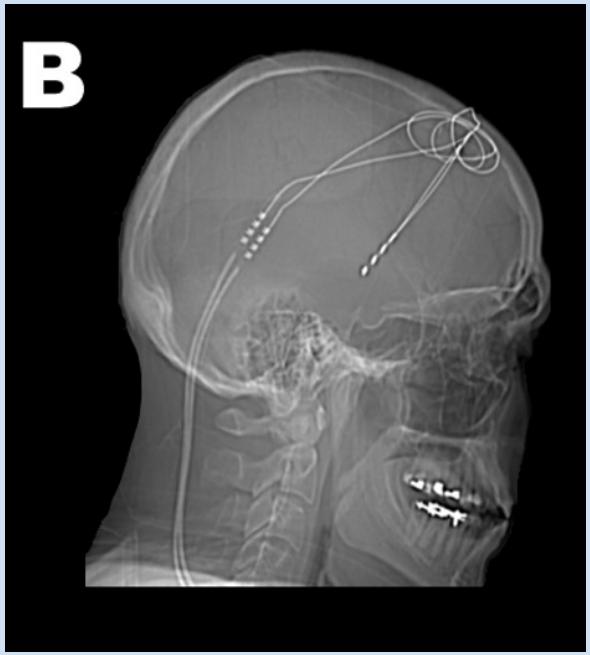
PeerJ

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

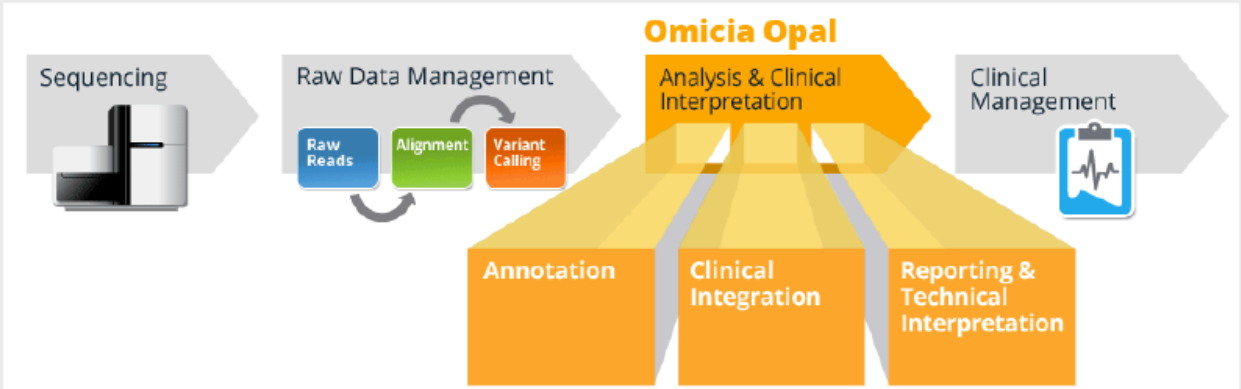
Additional Information and Declarations can be found on page 18

DOI **10.7717/peerj.177**

Commercial analysis platforms for genomic data

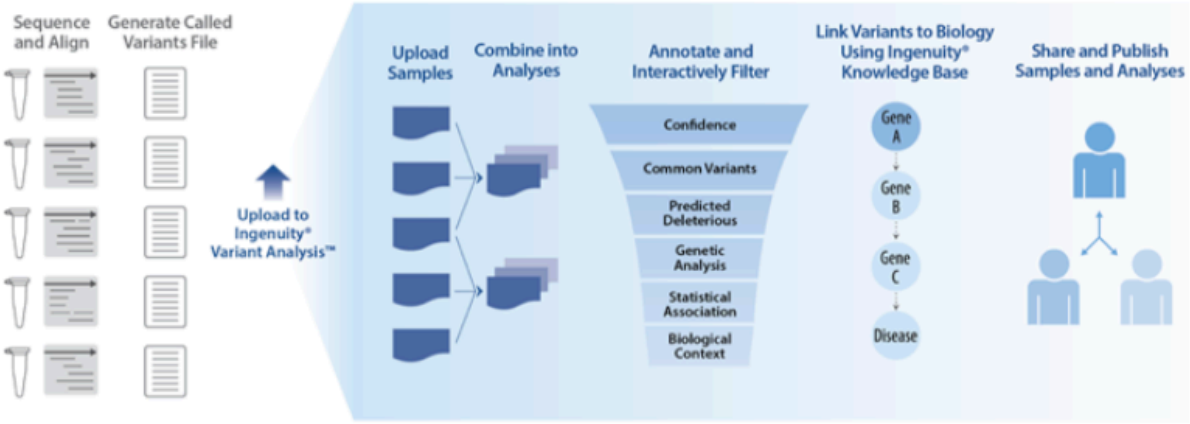
Opal adds clinical context for genomic data

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

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

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.