Ethics Panel: Challenges of Clinical Implementation of Genomic Medicine

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

STANLEY INSTITUTE FOR COGNITIVE GENOMICS
COLD SPRING HARBOR LABORATORY

UTAH FOUNDATION FOR BIOMEDICAL RESEARCH

@GholsonLyon

Down Syndrome

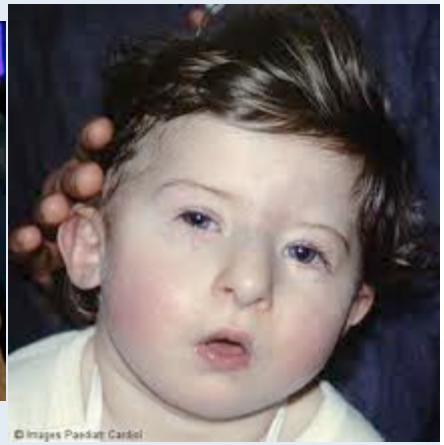
ARTICLE

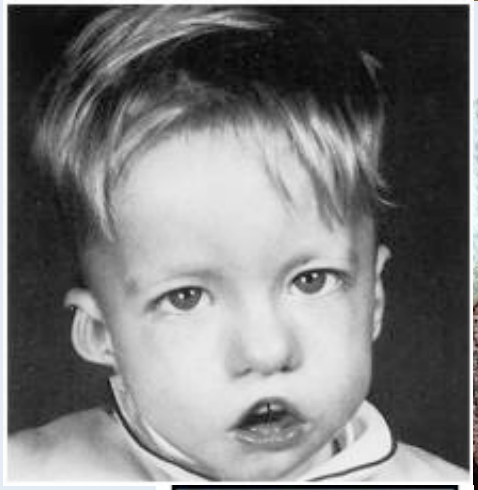
An Excess of Deleterious Variants in VEGF-A Pathway Genes in Down-Syndrome-Associated Atrioventricular Septal Defects

Christine Ackerman,[1] Adam E. Locke,[2,8] Eleanor Feingold,[3] Benjamin Reshey,[1] Karina Espana,[1] Janita Thusberg,[4] Sean Mooney,[4] Lora J.H. Bean,[2] Kenneth J. Dooley,[5] Clifford L. Cua,[6] Roger H. Reeves,[7] Stephanie L. Sherman,[2] and Cheryl L. Maslen[1,*]

About half of people with trisomy 21 have a congenital heart defect (CHD), whereas the remainder have a structurally normal heart, demonstrating that trisomy 21 is a significant risk factor but is not causal for abnormal heart development. Atrioventricular septal defects (AVSD) are the most commonly occurring heart defects in Down syndrome (DS), and ~65% of all AVSD is associated with DS. We used a candidate-gene approach among individuals with DS and complete AVSD (cases = 141) and DS with no CHD (controls = 141) to determine whether rare genetic variants in genes involved in atrioventricular valvuloseptal morphogenesis contribute to AVSD in this sensitized population. We found a significant excess ($p < 0.0001$) of variants predicted to be deleterious in cases compared to controls. At the most stringent level of filtering, we found potentially damaging variants in nearly 20% of cases but fewer than 3% of controls. The variants with the highest probability of being damaging in cases only were found in six genes: *COL6A1*, *COL6A2*, *CRELD1*, *FBLN2*, *FRZB*, and *GATA5*. Several of the case-specific variants were recurrent in unrelated individuals, occurring in 10% of cases studied. No variants with an equal probability of being damaging were found in controls, demonstrating a highly specific association with AVSD. Of note, all of these genes are in the VEGF-A pathway, even though the candidate genes analyzed in this study represented numerous biochemical and developmental pathways, suggesting that rare variants in the VEGF-A pathway might contribute to the genetic underpinnings of AVSD in humans.

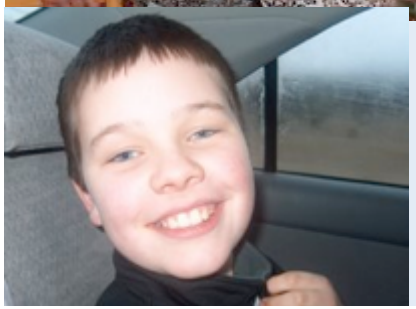
Velocardiofacial (22q11.2) Syndrome



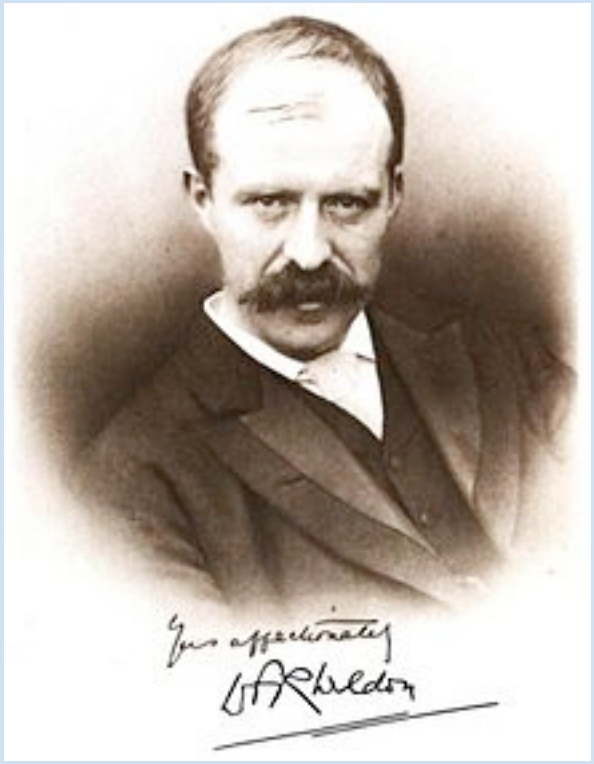




Beyond our Kuhnian inheritance

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.

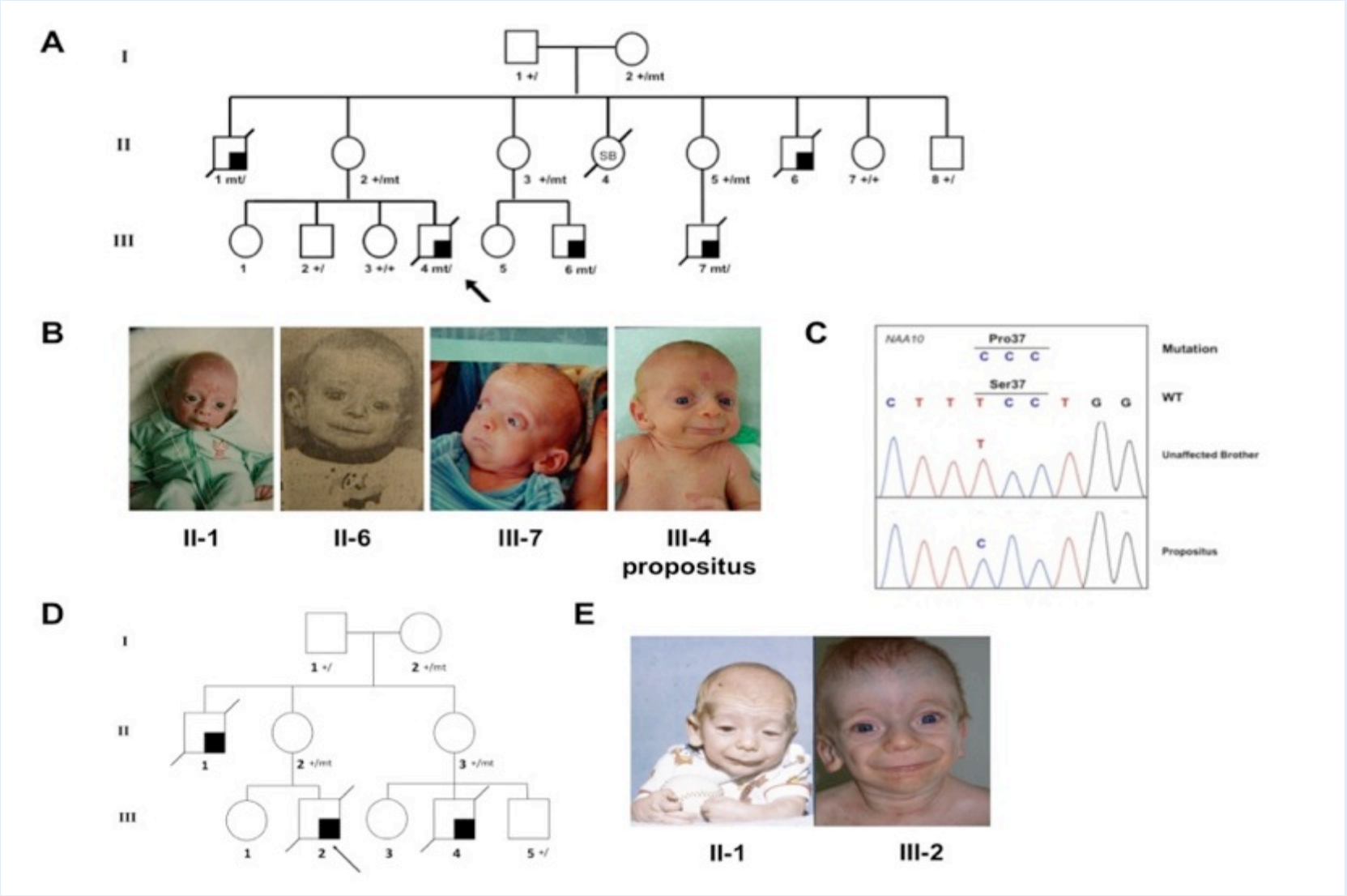
Plate I.

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

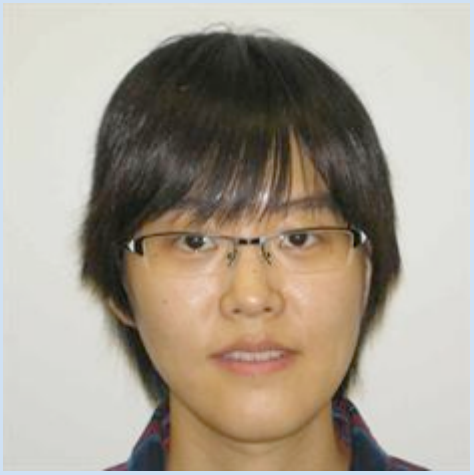
Penetrance and Expressivity

- We do not really know the penetrance or expressivity of pretty much ALL mutations in **humans**, as we have not systematically sequenced or karyotyped any genetic alteration in **MILLIONS** of well-phenotyped people.
- Do single mutations drive outcome predominately, or are the results modified substantially by other mutations and/or environment? Is there really such a thing as genetic determinism for **MANY** mutations?

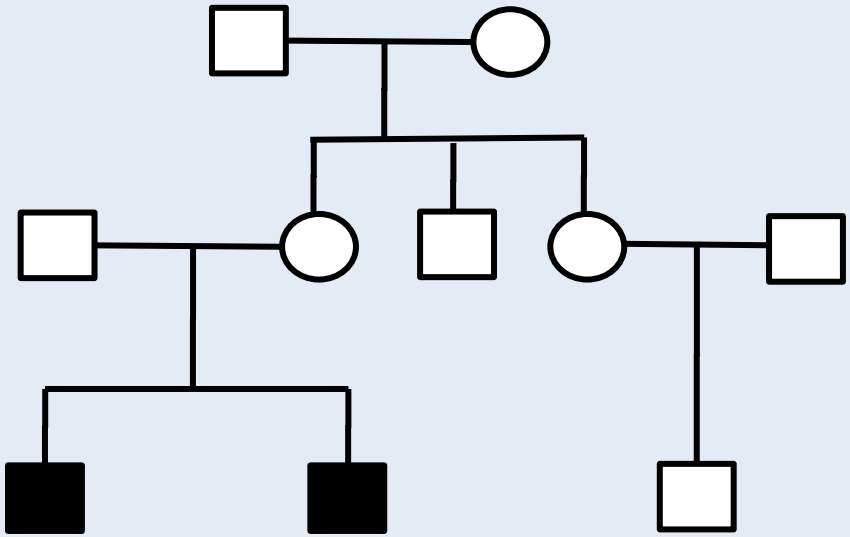
Ogden Syndrome



The mutation is **necessary**, but we do not know if it is **sufficient** to cause this phenotype in ANY genetic background. It simply “contributes to” the phenotype.



New Syndrome with Dysmorphology, 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

7 years old

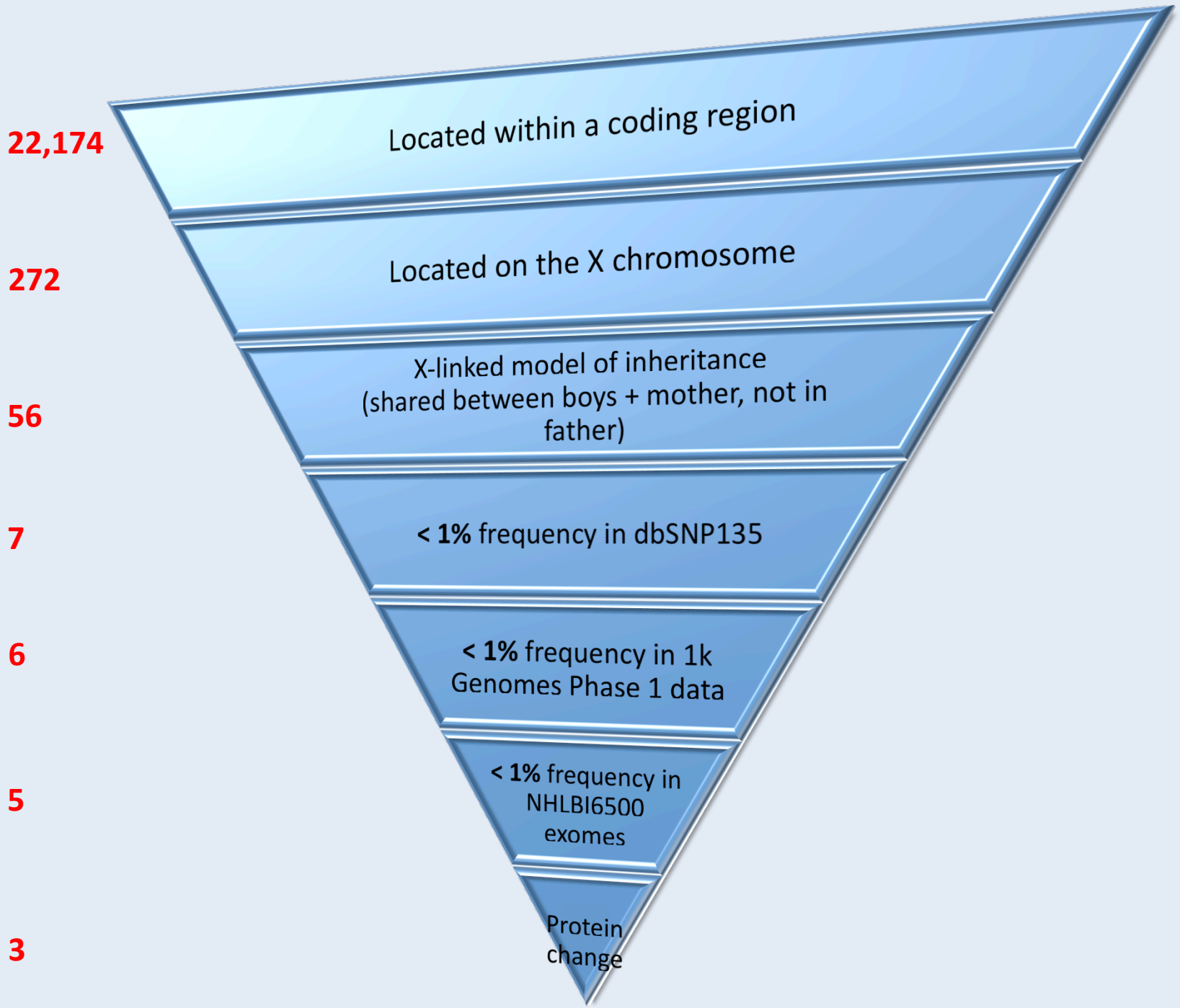
3 years old

5 years old

9 years old

Workup Ongoing for past 10 years

- Numerous genetic tests negative, including negative for Fragile X and many candidate genes.
- No obvious pathogenic CNVs – microarrays normal.
- Sequenced whole genomes of Mother, Father and Two Boys, using Complete Genomics, obtained data in June of this year, i.e. version 2.0 CG pipeline.


22,174
Located within a coding region
272
Located on the X chromosome
56
X-linked model of inheritance (shared between boys + mother, not in father)
7
< 1% frequency in dbSNP135
6
< 1% frequency in 1k Genomes Phase 1 data
5
< 1% frequency in NHLBI6500 exomes
3
Protein change

Variant classification

Variant	Reference	Alternate	Classification	Gene 1	Transcript 1	Exon 1	HGVS Coding 1	HGVS Protein 1
X:47307978-SNV	G	T	Nonsyn SNV	ZNF41	NM_007130	5	c.1191C>A	p.Asp397Glu
X:63444792-SNV	C	A	Nonsyn SNV	ASB12	NM_130388	2	c.739G>T	p.Gly247Cys
X:70621541-SNV	T	C	Nonsyn SNV	TAF1	NM_004606	25	c.4010T>C	p.Ile1337Thr

SIFT classification

Chromosome	Position	Reference	Coding?	SIFT Score	Score <= 0.05	Ref/Alt Alleles
X	47307978	G	YES	0.649999976	0	G/T
X	63444792	C	YES	0	1	C/A
X	70621541	T	YES	0.009999999776	1	T/C

VAAST score

RANK	Gene	p-value	p-value-ci	Score	Variants
1	ASB12	1.56E-11	1.55557809307134e-11,0.000290464582480396	38.63056297	chrX:63444792;38.63;C->A;G->C;0,3
2	TAF1	1.56E-11	1.55557809307134e-11,0.000290464582480396	34.51696816	chrX:70621541;34.52;T->C;I->T;0,3
3	ZNF41	1.56E-11	1.55557809307134e-11,0.000290464582480396	32.83011803	chrX:47307978;32.83;G->T;D->E;0,3

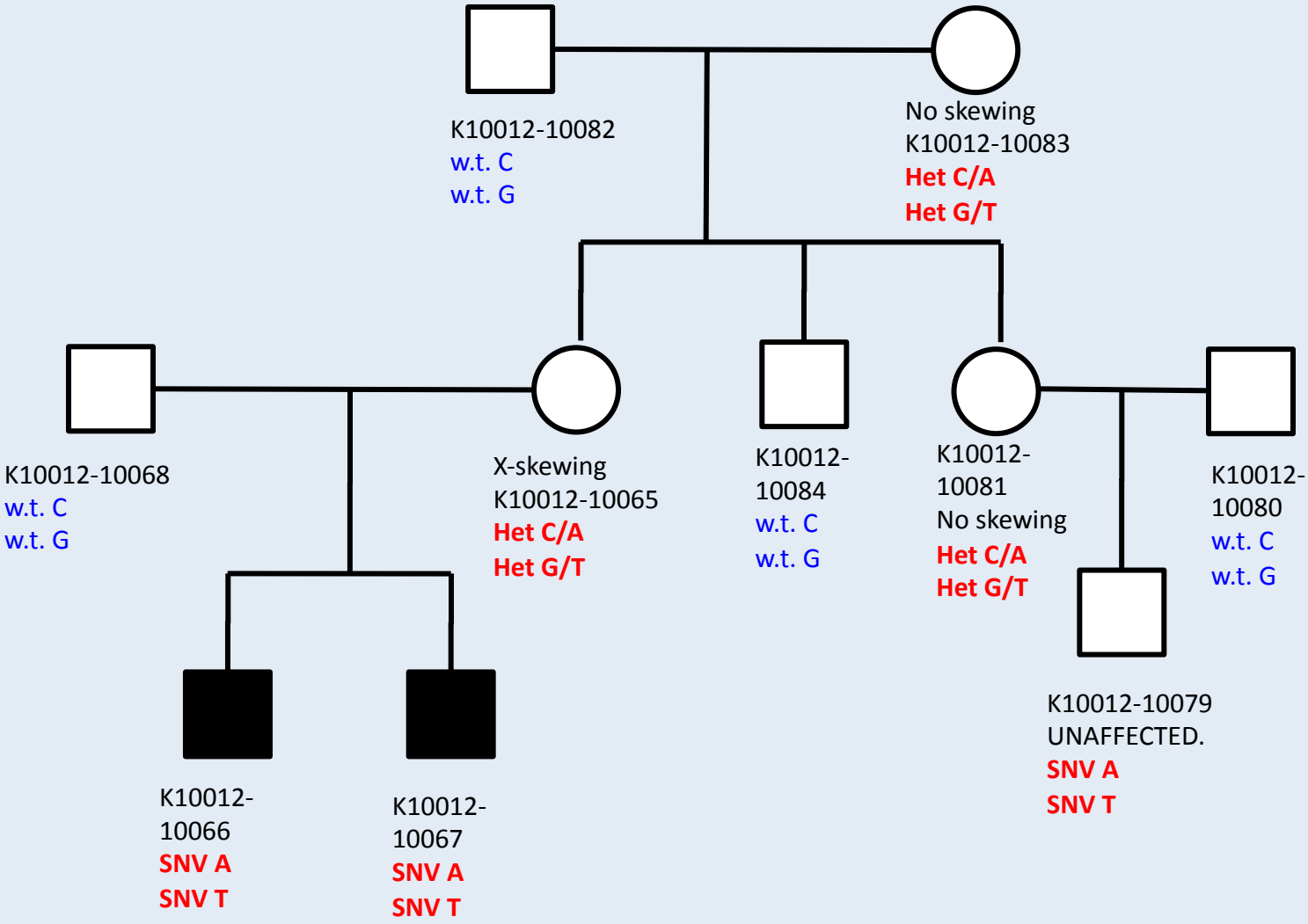
Mutations in the *ZNF41* Gene Are Associated with Cognitive Deficits: Identification of a New Candidate for X-Linked Mental Retardation

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

Sanger validation: ASB12 and ZNF41 mutations



The mutation in ZNF41 may **NOT** be necessary, and it is certainly **NOT** sufficient to cause the phenotype.

So, of course we need baseline whole genome sequencing on everyone to at least understand the DNA genetic background in each pedigree or clan.

Ancestry Matters!

How do we get to “whole” genome sequencing for everyone?

- ~$3000 for 30x whole genome as part of Illumina Genome Network on a research basis only, but ~$9,500 for whole genome performed in a CLIA lab at Illumina.
- The price point of $5,000 that was given for the Illumina Understand Your Genome Symposium was basically below cost (52 genomes delivered).
- And it is my understanding that many projects don’t want to pay even $5K for a whole genome in the CLIA lab at Illumina.

Method

Accurate and comprehensive sequencing of personal genomes

Subramanian S. Ajay,[1] Stephen C.J. Parker,[1] Hatice Ozel Abaan,[1] Karin V. Fuentes Fajardo,[2] and Elliott H. Margulies[1,3,4]

[1]*Genome Informatics Section, Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA;* [2]*Undiagnosed Diseases Program, Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA*

As whole-genome sequencing becomes commoditized and we begin to sequence and analyze personal genomes for clinical and diagnostic purposes, it is necessary to understand what constitutes a complete sequencing experiment for determining genotypes and detecting single-nucleotide variants. Here, we show that the current recommendation of ~30× coverage is not adequate to produce genotype calls across a large fraction of the genome with acceptably low error rates. Our results are based on analyses of a clinical sample sequenced on two related Illumina platforms, $GAII_x$ and HiSeq 2000, to a very high depth (126×). We used these data to establish genotype-calling filters that dramatically increase accuracy. We also empirically determined how the callable portion of the genome varies as a function of the amount of sequence data used. These results help provide a "sequencing guide" for future whole-genome sequencing decisions and metrics by which coverage statistics should be reported.

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

Complete Genomics – LFR technology

Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells

Brock A. Peters[1]*, Bahram G. Kermani[1]*, Andrew B. Sparks[1]†, Oleg Alferov[1], Peter Hong[1], Andrei Alexeev[1], Yuan Jiang[1], Fredrik Dahl[1]†, Y. Tom Tang[1], Juergen Haas[1], Kimberly Robasky[2,3], Alexander Wait Zaranek[2], Je-Hyuk Lee[2,4], Madeleine Price Ball[2], Joseph E. Peterson[1], Helena Perazich[1], George Yeung[1], Jia Liu[1], Linsu Chen[1], Michael I. Kennemer[1], Kaliprasad Pothuraju[1], Karel Konvicka[1], Mike Tsoupko-Sitnikov[1], Krishna P. Pant[1], Jessica C. Ebert[1], Geoffrey B. Nilsen[1], Jonathan Baccash[1], Aaron L. Halpern[1], George M. Church[2] & Radoje Drmanac[1]

NATURE | VOL 487 | 12 JULY 2012

"Substantial error rates (1 single nucleotide variants (SNV) in 100–1,000 called kilobases) are a common attribute of all current massively parallelized sequencing technologies. These rates are probably too high for diagnostic use and complicate many studies searching for new mutations."

Much Higher Accuracy with LFR data

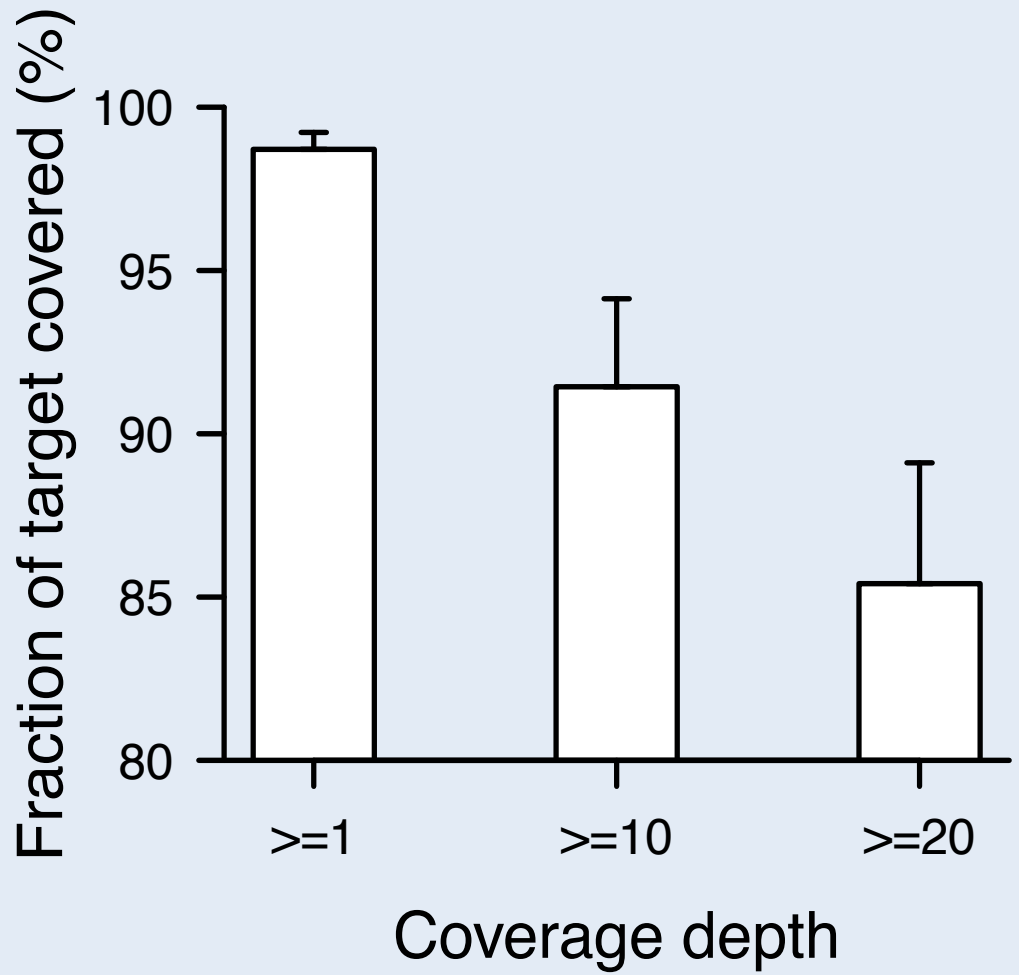
"To test LFR reproducibility we compared haplotype data between the two NA19240 replicate libraries. In general, the libraries were very concordant, with only 64 differences per library in 2.2 million heterozygous SNPs phased by both libraries or **1 of this error type in 44 Mb**."

Toward more comprehensive "personal genomes"

- Can we reliably detect a comprehensive, and accurate, set of variants using more than one pipeline, or even more than one sequencing platform?

- How much data is enough, and how reliable and reproducible are variant calls?

Depth of Coverage in 15 exomes > 20 reads per bp in target region



Pipelines Used on Same Set of Seq Data by Different Analysts, using Hg19 Reference Genome

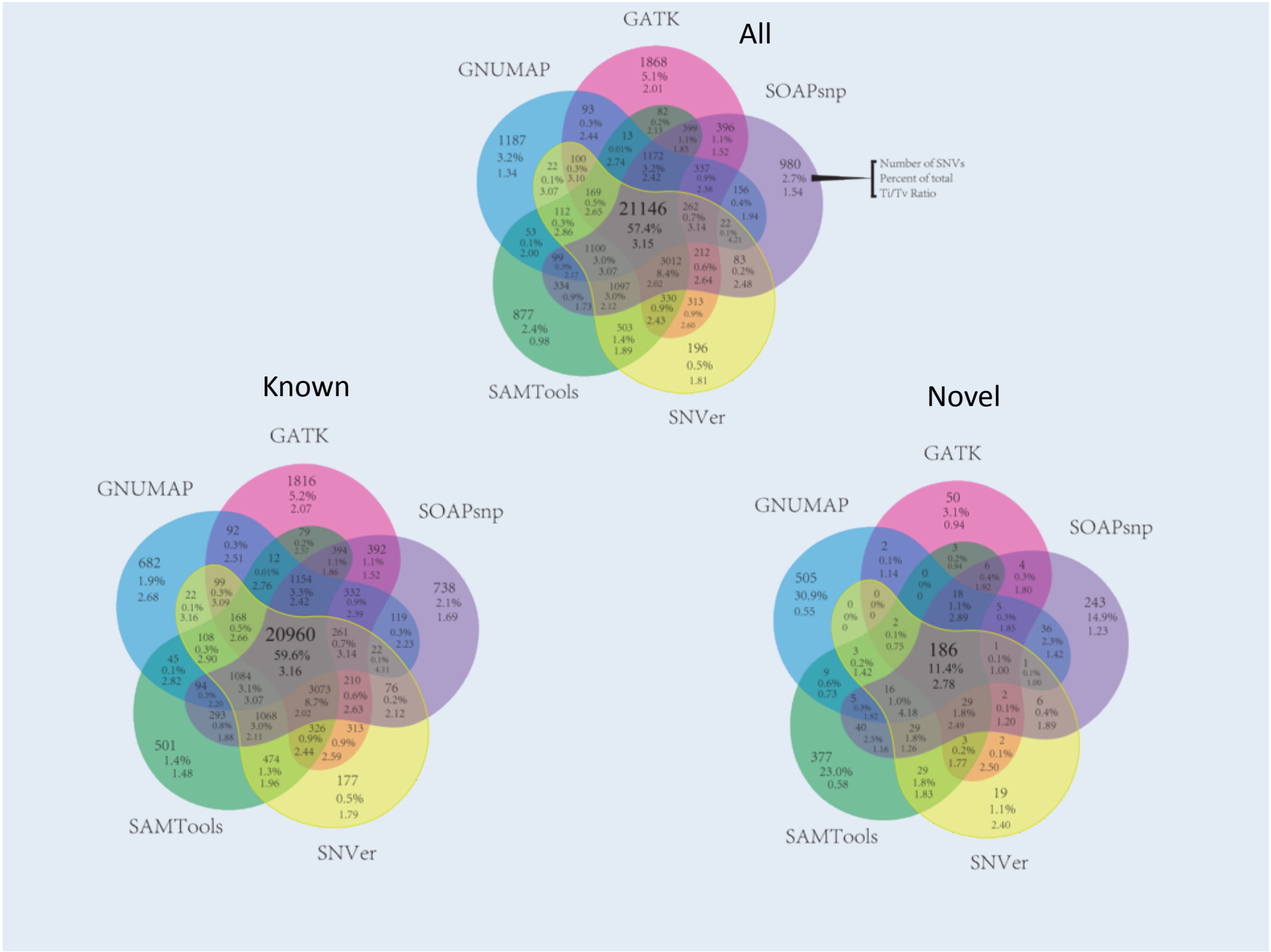
1) BWA - **GATK** (version 1.5) with recommended parameters (GATK IndelRealigner, base quality scores were re-calibrated by GATK Table Recalibration tool. Genotypes called by GATK UnifiedGenotyper. For SNVs and indels.

2) BWA - **SamTools** version 0.1.18 to generate genotype calls -- The "mpileup" command in SamTools was used for identify SNVs and indels.

3) SOAP-Align – SOAPsnp for SNVs– and BWA-SOAPindel (adopts local assembly based on an extended de Bruijn graph) for indels.

4) GNUMAP-SNP (probabilistic Pair-Hidden Markov which effectively accounts for uncertainty in the read calls as well as read mapping in an unbiased fashion), for SNVs only.

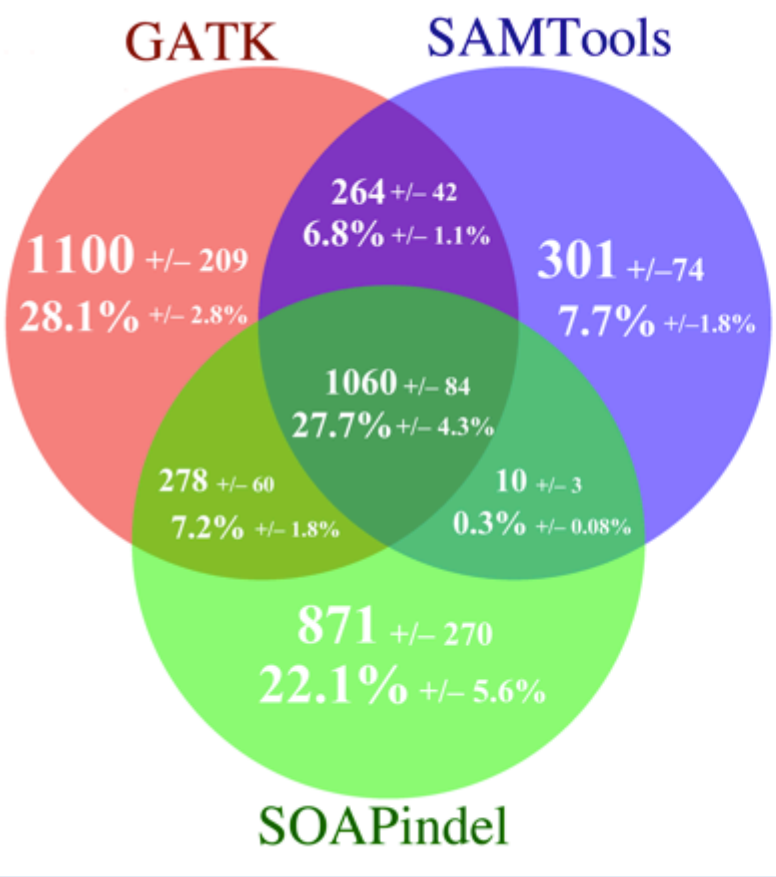
5) BWA - Sam format to Bam format - Picard to remove duplicates **– SNVer** , for SNVs only


All
GATK
GNUMAP
SOAPsnp
SAMTools
SNVer
1868
5.1%
2.01
1187
3.2%
1.34
980
2.7%
1.54
21146
57.4%
3.15
Number of SNVs
Percent of total
Ti/Tv Ratio
Known
GATK
GNUMAP
SOAPsnp
SAMTools
SNVer
1816
5.2%
2.07
682
1.9%
2.68
738
2.1%
1.69
20960
59.6%
3.16
501
1.4%
1.48
177
0.5%
1.79
Novel
GATK
GNUMAP
SOAPsnp
SAMTools
SNVer
50
3.1%
0.94
505
30.9%
0.55
243
14.9%
1.23
186
11.4%
2.78
377
23.0%
0.58
19
1.1%
2.40

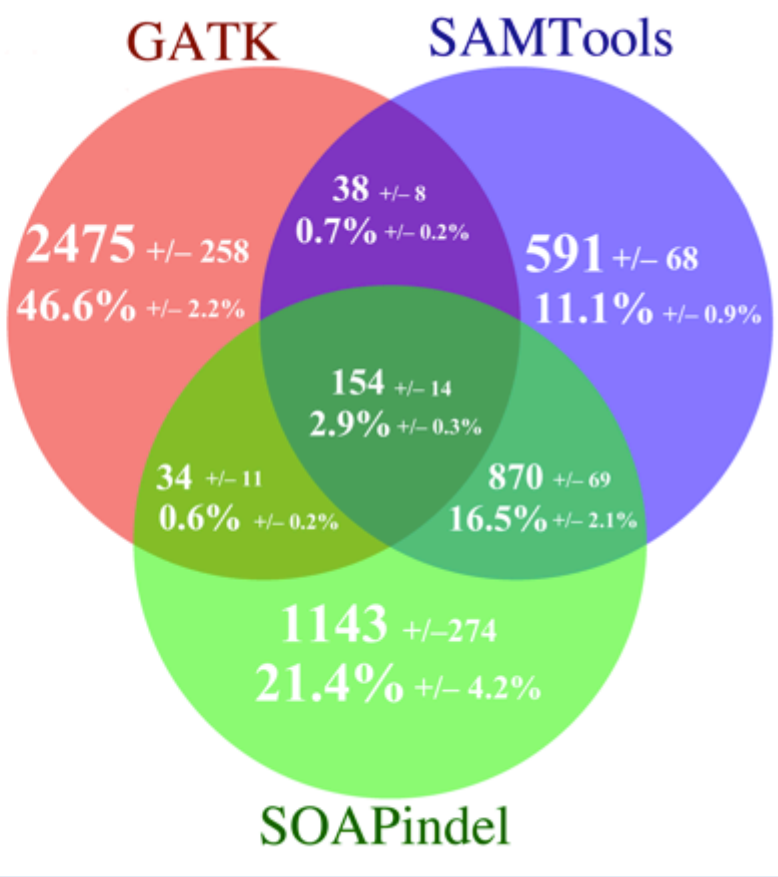
- Pipeline concordance is much better for known and common human variations (in dbSNP135).
- How reliable are variants that are uniquely called by individual pipelines?
- Are some pipelines better at detecting rare, or novel variants than others?

INDELS

Indels- Overlap by Base Position only

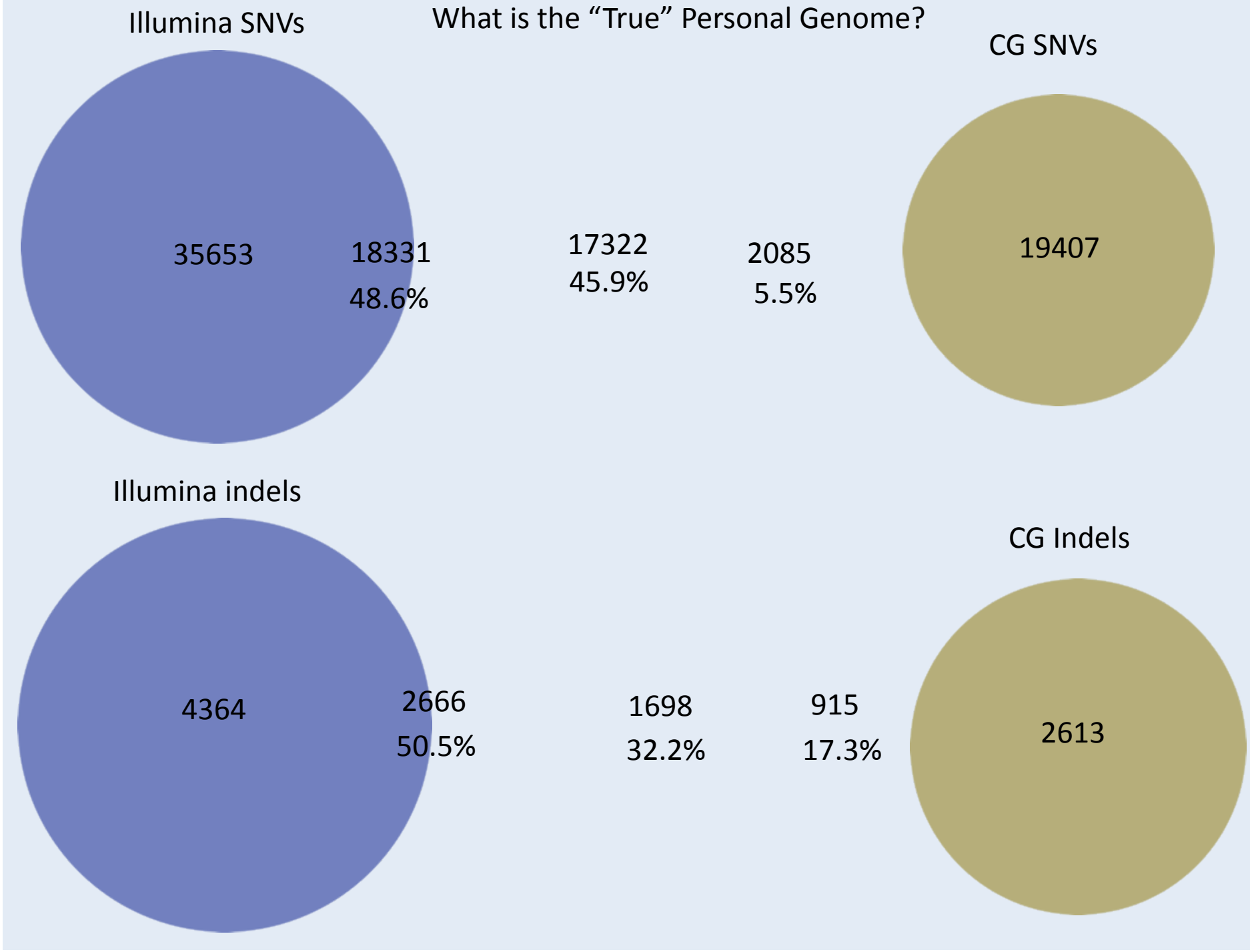


Indels- Overlap by Base Position, Length **and** Composition

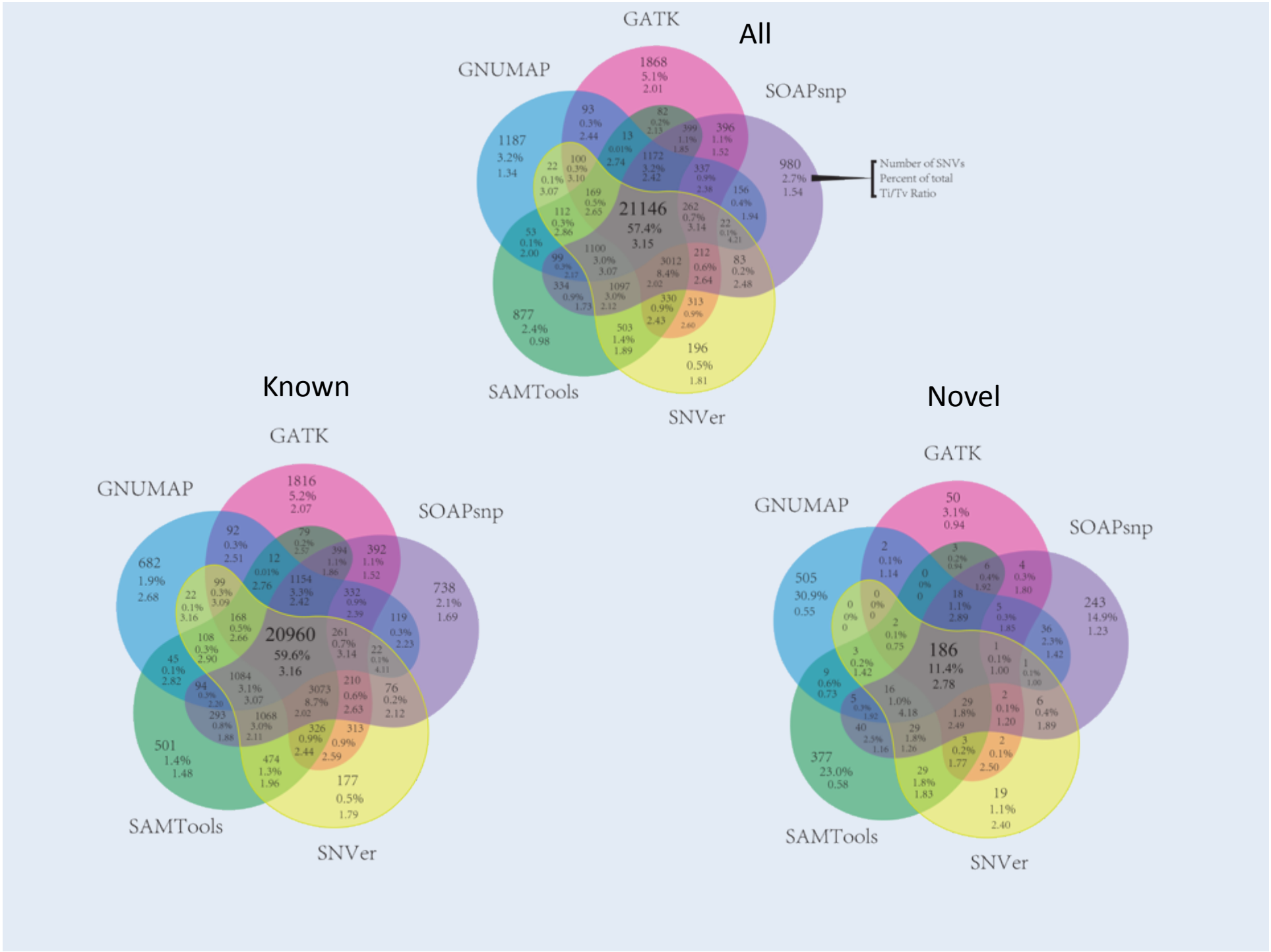


Total mean overlap, plus or minus one standard deviation, observed between three indel calling pipelines: GATK, SOAP-indel, and SAMTools. a) Mean overlap when indel position was the only necessary agreement criterion. **b)** Mean overlap when indel position, base length and base composition were the necessary agreement criteria.

What is the “True” Personal Genome?

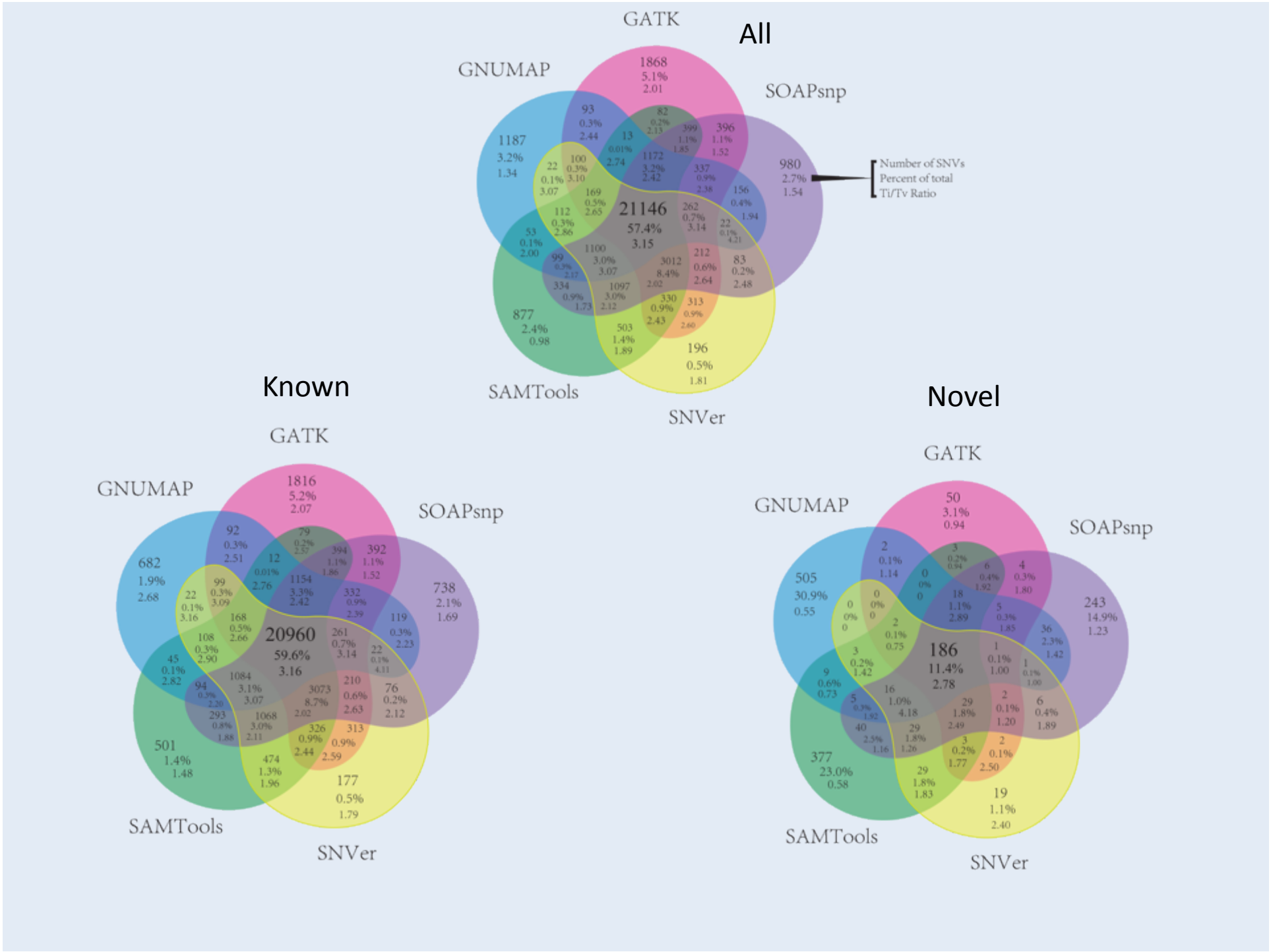


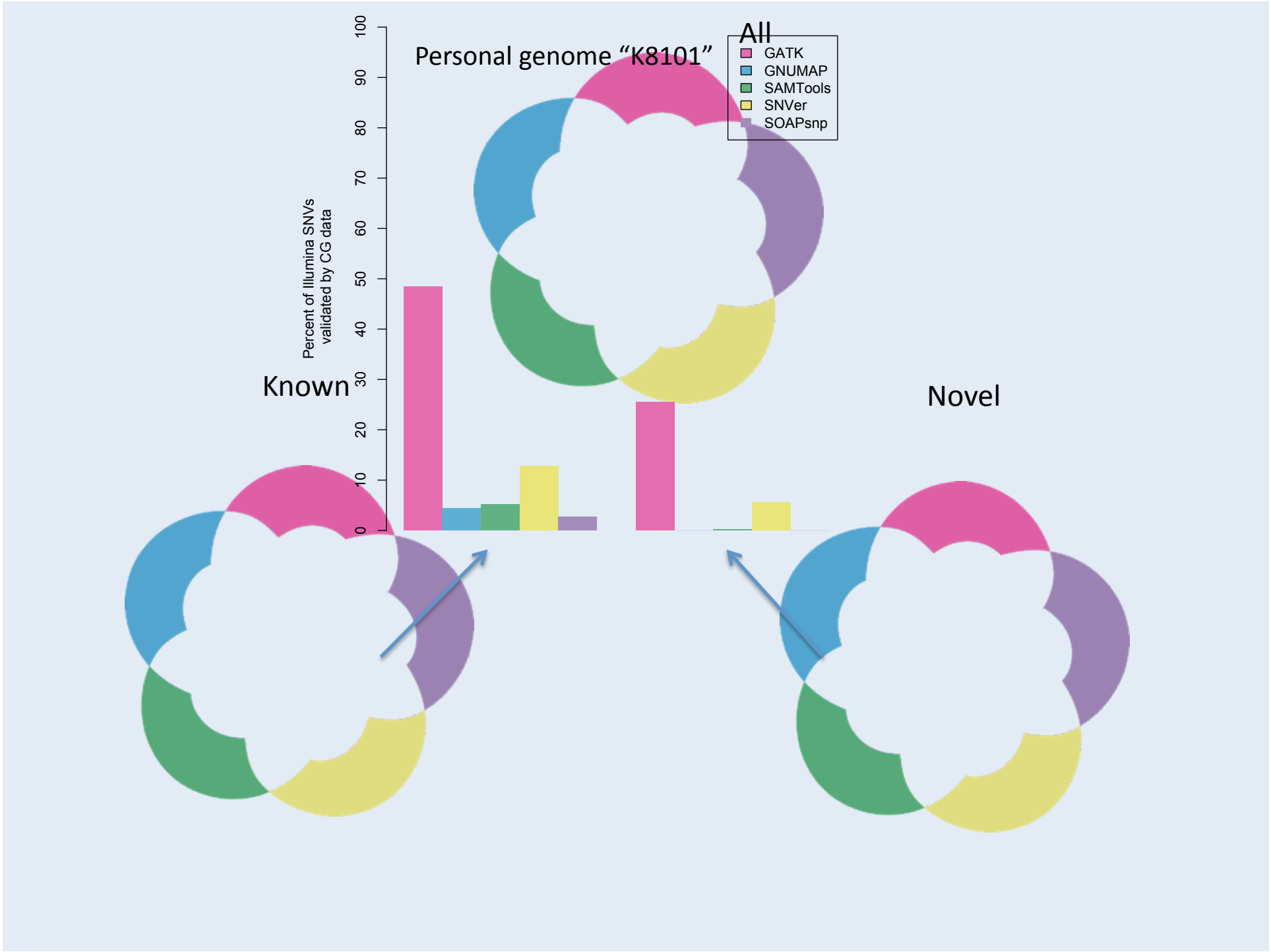
Cross validation using orthogonal sequencing technology (Complete Genomics)


All
GATK
GNUMAP
SOAPsnp
SAMTools
SNVer
1868
5.1%
2.01
1187
3.2%
1.34
980
2.7%
1.54
21146
57.4%
3.15
Number of SNVs
Percent of total
Ti/Tv Ratio
Known
GATK
GNUMAP
SOAPsnp
SAMTools
SNVer
1816
5.2%
2.07
682
1.9%
2.68
738
2.1%
1.69
20960
59.6%
3.16
501
1.4%
1.48
177
0.5%
1.79
Novel
GATK
GNUMAP
SOAPsnp
SAMTools
SNVer
50
3.1%
0.94
505
30.9%
0.55
243
14.9%
1.23
186
11.4%
2.78
377
23.0%
0.58
19
1.1%
2.40

Cross validation using orthogonal sequencing tech (Complete Genomics)

- Reveals higher validation rate of unique-to-pipeline variants, as well as uniquely discovered novel variants, for the variants called by BWA-GATK, in comparison to the other 4 pipelines (including SOAP).


All
GATK
GNUMAP
SOAPsnp
SAMTools
SNVer
1868
5.1%
2.01
1187
3.2%
1.34
980
2.7%
1.54
21146
57.4%
3.15
Number of SNVs
Percent of total
Ti/Tv Ratio
Known
GATK
GNUMAP
SOAPsnp
SAMTools
SNVer
1816
5.2%
2.07
682
1.9%
2.68
738
2.1%
1.69
20960
59.6%
3.16
501
1.4%
1.48
177
0.5%
1.79
Novel
GATK
GNUMAP
SOAPsnp
SAMTools
SNVer
50
3.1%
0.94
505
30.9%
0.55
243
14.9%
1.23
186
11.4%
2.78
377
23.0%
0.58
19
1.1%
2.40


Personal genome “K8101”
All
GATK
GNUMAP
SAMTools
SNVer
SOAPsnp
Percent of Illumina SNVs validated by CG data
0
10
20
30
40
50
60
70
80
90
100
Known
Novel

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.

Clinical Validity with Worldwide Human Genetic Variation "database"?

PatientsLikeMe

PRIVACY and PROGRESS in Whole Genome Sequencing

Presidential Commission
for the Study of Bioethical Issues

October 2012

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]

Conclusions

- Ancestry, i.e. genetic background, matters!
- We need to sequence whole genomes of large pedigrees, and then construct super-family structures.
- 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 consumers.

Acknowledgments

Jason O'Rawe
Yiyang Wu
Michael Schatz
Giuseppe Narzisi

Reid Robison
Edwin Nyambi

USC
Kai Wang

Tao Jiang
Guangqing Sun
Jun Wang

our study families

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]